NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Computer Engineering

Interfacing Between PC and Mobile using Bluetooth

Graduation Project
COM- 400

Student :
YOUSEF AL – SAKARNEH (20011166)

Supervisor : Mr. Jamal Fathi

Nicosia-2005

ACKNOWLEDGMENT

"First , i would like to thank my supervisor Mr Jamal Fathi for his invaluable and belif in my work and my self over the course of this graduation project.

Second , i thank my father and, my brothers and my sister for their constant encouragement and support during my study and preparation of this project.

Finally , i would also like to thank all my friends for their help and support."

ABSTRACT

The mobile station (MS) consists of the mobile equipment (the terminal) and a smart card called the Subscriber Identity Module (SIM). The SIM provides personal mobility, so that the user can have access to subscribed services irrespective of a specific terminal. By inserting the SIM card into another GSM terminal, the user is able to receive calls at that terminal, make calls from that terminal, and receive other subscribed services.

The GSM technical specifications define the different entities that form the GSM network by defining their functions and interface requirements.

Each mobile uses a separate, temporary radio channel to talk to the cell site. The cell site talks to many mobiles at once, using one channel per mobile. Channels use a pair of frequencies for communication—one frequency (the forward link) for transmitting from the cell site and one frequency (the reverse link) for the cell site to receive calls from the users. Radio energy dissipates over distance, so mobiles must stay near the base station to maintain communications. The basic structure of mobile networks includes telephone systems and radio services. Where mobile radio service operates in a closed network and has no access to the telephone system, mobile telephone service allows interconnection to the telephone network.

CONTENTS

1. THE GSM NETWORK

1.1 Overview

During the early 1980s, analog cellular telephone systems were experiencing rapid growth in Europe, particularly in Scandinavia and the United Kingdom, but also in France and Germany. Each country developed its own system, which was incompatible with everyone else's in equipment and operation. This was an undesirable situation, because not only was the mobile equipment limited to operation within national boundaries, which in a unified Europe were increasingly unimportant, but there was also a very limited market for each type of equipment, so economies of scale and the subsequent savings could not be realized.

The Europeans realized this early on, and in 1982 the Conference of European Posts and Telegraphs (CEPT) formed a study group called the Group Special Mobile (GSM) to study and develop a pan-European public land mobile system. The proposed system had to meet certain criteria:

1. Good subjective speech quality
2. Low terminal and service cost
3. Support for international roaming
4. Ability to support handheld terminals
5. Support for range of new services and facilities
6. Spectral efficiency
7. ISDN compatibility

In 1989, GSM responsibility was transferred to the European Telecommunication Standards Institute (ETSI), and phase I of the GSM specifications were published in 1990. Commercial service was started in mid-1991, and by 1993 there were 36 GSM networks in 22 countries. Although standardized in Europe, GSM is not only a European standard. Over 200 GSM networks (including DCS1800 and PCS1900) are operational in 110 countries around the world. In the beginning of 1994, there were 1.3 million subscribers worldwide, which had grown to more than 55 million by October 1997.

With North America making a delayed entry into the GSM field with a derivative of GSM called PCS1900, GSM systems exist on every continent, and the acronym GSM now aptly stands for Global System for Mobile communications.

The developers of GSM chose an unproven (at the time) digital system, as opposed to the then-standard analog cellular systems like AMPS in the United States and TACS in the United Kingdom. They had faith that advancements in compression algorithms and digital signal processors would allow the fulfillment of the original criteria and the continual improvement of the system in terms of quality and cost. The over 8000 pages of GSM recommendations try to allow flexibility and competitive innovation among suppliers, but provide enough standardization to guarantee proper inter working between the components of the system. This is done by providing functional and interface descriptions for each of the functional entities defined in the system.

1.2 Services provided by GSM

From the beginning, the planners of GSM wanted ISDN compatibility in terms of the services offered and the control signaling used. However, radio transmission limitations, in terms of bandwidth and cost, do not allow the standard ISDN B-channel bit rate of 64 kbps to be practically achieved.

Using the ITU-T definitions, telecommunication services can be divided into bearer services, tele-services, and supplementary services. The most basic tele-service supported by GSM is telephony. As with all other communications, speech is digitally encoded and transmitted through the GSM network as a digital stream. There is also an emergency service, where the nearest emergency-service provider is notified by dialing three digits (similar to 911).

A variety of data services is offered. GSM users can send and receive data, at rates up to 9600 bps, to users on POTS (Plain Old Telephone Service), ISDN, Packet Switched Public Data Networks, and Circuit Switched Public Data Networks using a variety of access methods and protocols, such as X.25 or X.32. Since GSM is a digital network, a modem is not required between the user and GSM network, although an audio modem is required inside the GSM network to inter work with POTS.

Other data services include Group 3 facsimile, as described in ITU-T recommendation T.30, which is supported by use of an appropriate fax adaptor. A unique feature of GSM, not found in older analog systems, is the Short Message Service (SMS). SMS is a bi directional service for short alphanumeric (up to 160 bytes) messages. Messages are transported in a store-and-forward fashion. For point-to-point SMS, a message can be sent to another subscriber to the service, and an acknowledgement of receipt is provided to the sender. SMS can also be used in a cell-broadcast mode, for sending messages such as traffic updates or news updates. Messages can also be stored in the SIM card for later retrieval.

Supplementary services are provided on top of tele-services or bearer services. In the current (Phase I) specifications, they include several forms of call forward (such as call forwarding when the mobile subscriber is unreachable by the network), and call barring of outgoing or incoming calls, for example when roaming in another country. Many additional supplementary services will be provided in the Phase 2 specifications, such as caller identification, call waiting, multi-party conversations.

1.3 Architecture of the GSM network

A GSM network is composed of several functional entities, whose functions and interfaces are specified. Figure 1.1 shows the layout of a generic GSM network. The GSM network can be divided into three broad parts. The Mobile Station is carried by the subscriber. The Base Station Subsystem controls the radio link with the Mobile Station. The Network Subsystem, the main part of which is the Mobile services Switching Center (MSC), performs the switching of calls between the mobile users, and between mobile and fixed network users. The MSC also handles the mobility management operations. Not shown is the Operations and Maintenance Center, which oversees the proper operation and setup of the network. The Mobile Station and the Base Station Subsystem communicate across the Um interface, also known as the air interface or radio link. The Base Station Subsystem communicates with the Mobile services Switching Center across the A interface.

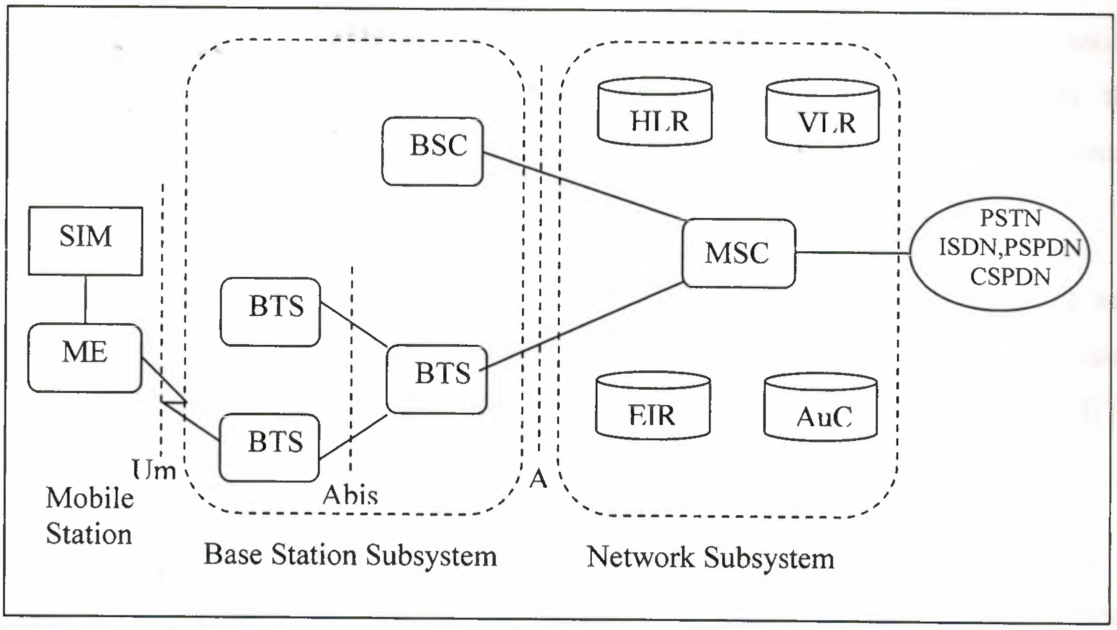

Figure 1.1 General Architecture of a GSM Network

1.3.1 Mobile Station

The mobile station (MS) consists of the mobile equipment (the terminal) and a smart card called the Subscriber Identity Module (SIM). The SIM provides personal mobility, so that the user can have access to subscribed services irrespective of a specific terminal. By inserting the SIM card into another GSM terminal, the user is able to receive calls at that terminal, make calls from that terminal, and receive other subscribed services.

The mobile equipment is uniquely identified by the International Mobile Equipment Identity (IMEI). The SIM card contains the International Mobile Subscriber Identity (IMSI) used to identify the subscriber to the system, a secret key for authentication, and other information. The IMEI and the IMSI are independent, thereby allowing personal mobility. The SIM card may be protected against unauthorized use by a password or personal identity number.

1.3.2 Base Station Subsystem

The Base Station Subsystem is composed of two parts, the Base Transceiver Station (BTS) and the Base Station Controller (BSC). These communicate across the standardized Abis interface, allowing (as in the rest of the system) operation between components made by different suppliers.

The Base Transceiver Station houses the radio transceivers that define a cell and handles the radio-link protocols with the Mobile Station. In a large urban area, there will potentially be a large number of BTSs deployed, thus the requirements for a BTS are ruggedness, reliability, portability, and minimum cost.

The Base Station Controller manages the radio resources for one or more BTSs. It handles radio-channel setup, frequency hopping, and handovers, as described below. The BSC is the connection between the mobile station and the Mobile service Switching Center (MSC).

1.3.3 Network Subsystem

The central component of the Network Subsystem is the Mobile services Switching Center (MSC). It acts like a normal switching node of the PSTN or ISDN, and additionally provides all the functionality needed to handle a mobile subscriber, such as registration, authentication, location updating, handovers, and call routing to a roaming subscriber. These services are provided in conjunction with several functional entities, which together form the Network Subsystem. The MSC provides the connection to the fixed networks (such as the PSTN or ISDN). Signaling between functional entities in the Network Subsystem uses Signaling System Number 7 (SS7), used for trunk signaling in ISDN and widely used in current public networks.

The Home Location Register (HLR) and Visitor Location Register (VLR), together with the MSC, provide the call-routing and roaming capabilities of GSM. The HLR contains all the administrative information of each subscriber registered in the corresponding GSM network, along with the current location of the mobile. The location of the mobile is typically in the form of the signaling address of the VLR associated with the mobile station. The actual routing procedure will be described later. There is logically one HLR per GSM network, although it may be implemented as a distributed database.

The Visitor Location Register (VLR) contains selected administrative information from the HLR, necessary for call control and provision of the subscribed services, for each mobile currently located in the geographical area controlled by the VLR. Although each functional entity can be implemented as an independent unit, all manufacturers of switching equipment to date implement the VLR together with the MSC, so that the geographical area controlled by the MSC corresponds to that controlled by the VLR, thus simplifying the signaling required. Note that the MSC contains no information about particular mobile stations --- this information is stored in the location registers.

The other two registers are used for authentication and security purposes. The Equipment Identity Register (EIR) is a database that contains a list of all valid mobile equipment on the network, where each mobile station is identified by its International Mobile Equipment Identity (IMEI). An IMEI is marked as invalid if it has been reported stolen or is not type approved. The Authentication Center (AuC) is a protected database that stores a copy of the secret key stored in each subscriber's SIM card, which is used for authentication and encryption over the radio channel.

1.4 Radio Link Aspects

The International Telecommunication Union (ITU), which manages the international allocation of radio spectrum (among many other functions), allocated the bands 890-915 MHz for the uplink (mobile station to base station) and 935-960 MHz for the downlink (base station to mobile station) for mobile networks in Europe. Since this range was already being used in the early 1980s by the analog systems of the day, the CEPT had the foresight to reserve the top 10 MHz of each band for the GSM network that was still being developed. Eventually, GSM will be allocated the entire 2x25 MHz bandwidth.

1.4.1 Multiple Access and Channel Structure

Since radio spectrum is a limited resource shared by all users, a method must be devised to divide up the bandwidth among as many users as possible. The method chosen by GSM is a combination of Time- and Frequency-Division Multiple Access (TDMA/FDMA). The FDMA part involves the division by frequency of the (maximum) 25 MHz bandwidth into 124 carrier frequencies spaced 200 kHz apart. One or more carrier frequencies are assigned to each base station. Each of these carrier frequencies is then divided in time, using a TDMA scheme. The fundamental unit of time in this TDMA scheme is called a *burst period* and it lasts 15/26 ms (or approx. 0.577 ms). Eight burst periods are grouped into a *TDMA frame* (120/26 ms, or approx. 4.615 ms), which forms the basic unit for the definition of logical channels. One physical channel is one burst period per TDMA frame.

Channels are defined by the number and position of their corresponding burst periods. All these definitions are cyclic, and the entire pattern repeats approximately every 3 hours. Channels can be divided into *dedicated channels*, which are allocated to a mobile station, and *common channels*, which are used by mobile stations in idle mode.

1.4.1.1 Traffic Channels

A traffic channel (TCH) is used to carry speech and data traffic. Traffic channels are defined using a 26-frame multi-frame, or group of 26 TDMA frames. The length of a 26-frame multi-frame is 120 ms, which is how the length of a burst period is defined (120 ms divided by 26 frames divided by 8 burst periods per frame). Out of the 26 frames, 24 are used for traffic, 1 is used for the Slow Associated Control Channel (SACCH) and 1 is currently unused (see Figure 1.2). TCHs for the uplink and downlink are separated in time by 3 burst periods, so that the mobile station does not have to transmit and receive simultaneously, thus simplifying the electronics.

In addition to these *full-rate* TCHs, there are also *half-rate* TCHs defined, although they are not yet implemented. Half-rate TCHs will effectively double the capacity of a system once half-rate speech coders are specified (i.e., speech coding at around 7 kbps, instead of 13 kbps). Eighth-rate TCHs are also specified, and are used for signaling. In the recommendations, they are called Stand-alone Dedicated Control Channels (SDCCH).

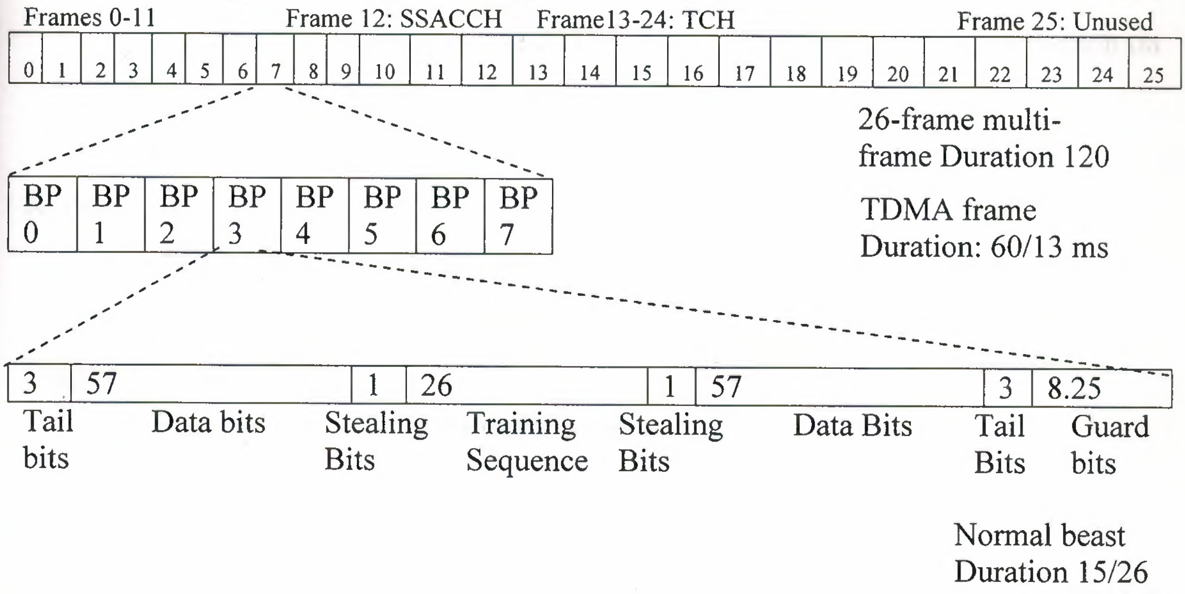

Figure 1.2. Organization of bursts, TDMA frames, and multi-frames for speech and data

1.4.1.2 Control Channels

Common channels can be accessed both by idle mode and dedicated mode mobiles. The common channels are used by idle mode mobiles to exchange the signaling information required to change to dedicated mode. Mobiles already in dedicated mode monitor the surrounding base stations for handover and other information. The common channels are defined within a 51-frame multi-frame, so that dedicated mobiles using the 26-frame multi-frame TCH structure can still monitor control channels. The common channels include:

Broadcast Control Channel (BCCH)

Continually broadcasts, on the downlink, information including base station identity, frequency allocations, and frequency-hopping sequences.

Frequency Correction Channel (FCCH) and Synchronization Channel (SCH)

Used to synchronize the mobile to the time slot structure of a cell by defining the boundaries of burst periods, and the time slot numbering. Every cell in a GSM network broadcasts exactly one FCCH and one SCH, which are by definition on time slot number 0 (within a TDMA frame).

Random Access Channel (RACH)

Slotted Aloha channel used by the mobile to request access to the network.

Paging Channel (PCH)

Used to alert the mobile station of an incoming call.

Access Grant Channel (AGCH)

Used to allocate an SDCCH to a mobile for signaling (in order to obtain a dedicated channel), following a request on the RACH.

1.4.1.3 Burst Structure

There are four different types of bursts used for transmission in GSM. The normal burst is used to carry data and most signaling. It has a total length of 156.25 bits, made up of two 57 bit information bits, a 26 bit training sequence used for equalization, 1 stealing bit for each information block (used for FACCH), 3 tail bits at each end, and an 8.25 bit guard sequence, as shown in Figure 1.2 The 156.25 bits are transmitted in 0.577 ms, giving a gross bit rate of 270.833 kbps.

The F burst, used on the FCCH, and the S burst, used on the SCH, have the same length as a normal burst, but a different internal structure, which differentiates them from normal bursts (thus allowing synchronization). The access burst is shorter than the normal burst, and is used only on the RACH.

1.4.2 Speech Coding

GSM is a digital system, so speech which is inherently analog, has to be digitized. The method employed by ISDN, and by current telephone systems for multiplexing voice lines over high speed trunks and optical fiber lines, is Pulse Coded Modulation (PCM). The output stream from PCM is 64 kbps, too high a rate to be feasible over a radio link. The 64 kbps signal, although simple to implement, contains much redundancy. The GSM group studied several speech coding algorithms on the basis of subjective speech quality and complexity (which is related to cost, processing delay, and power consumption once implemented) before arriving at the choice of a Regular Pulse Excited -- Linear Predictive Coder (RPE--LPC) with a Long Term Predictor loop. Basically, information from previous samples, which does not change very quickly, is used to predict the current sample. The coefficients of the linear combination of the previous samples, plus an encoded form of the residual, the difference between the predicted and actual sample, represent the signal. Speech is divided into 20 millisecond samples, each of which is encoded as 260 bits, giving a total bit rate of 13 kbps. This is the so-called Full-Rate speech coding. Recently, an Enhanced Full-Rate (EFR) speech coding algorithm has been implemented by some North American GSM1900 operators. This is said to provide improved speech quality using the existing 13 kbps bit rate.

1.4.3 Channel Coding and Modulation

Because of natural and man-made electromagnetic interference, the encoded speech or data signal transmitted over the radio interface must be protected from errors. GSM uses convolution encoding and block interleaving to achieve this protection. The exact algorithms used differ for speech and for different data rates. The method used for speech blocks will be described below.

Recall that the speech code produces a 260 bit block for every 20 ms speech sample. From subjective testing, it was found that some bits of this block were more important for perceived speech quality than others.

The bits are thus divided into three classes:

- **Class Ia** 50 bits - most sensitive to bit errors
- **Class Ib** 132 bits - moderately sensitive to bit errors
- **Class II** 78 bits - least sensitive to bit errors

Class Ia bits have a 3 bit Cyclic Redundancy Code added for error detection. If an error is detected, the frame is judged too damaged to be comprehensible and it is discarded. It is replaced by a slightly attenuated version of the previous correctly received frame. These 53 bits, together with the 132 Class Ib bits and a 4 bit tail sequence (a total of 189 bits), are input into a 1/2 rate convolution encoder of constraint length 4. Each input bit is encoded as two output bits, based on a combination of the previous 4 input bits. The convolution encoder thus outputs 378 bits, to which are added the 78 remaining Class II bits, which are unprotected. Thus every 20 ms speech sample is encoded as 456 bits, giving a bit rate of 22.8 kbps.

To further protect against the burst errors common to the radio interface, each sample is interleaved. The 456 bits output by the convolution encoder are divided into 8 blocks of 57 bits, and these blocks are transmitted in eight consecutive time-slot bursts. Since each time-slot burst can carry two 57 bit blocks, each burst carries traffic from two different speech samples.

Recall that each time-slot burst is transmitted at a gross bit rate of 270.833 kbps. This digital signal is modulated onto the analog carrier frequency using Gaussian-filtered Minimum Shift Keying (GMSK). GMSK was selected over other modulation schemes as a compromise between spectral efficiency, complexity of the transmitter, and limited spurious emissions. The complexity of the transmitter is related to power consumption, which should be minimized for the mobile station. The spurious radio emissions, outside of the allotted bandwidth, must be strictly controlled so as to limit adjacent channel interference, and allow for the co-existence of GSM and the older analog systems (at least for the time being).

1.4.4 Multipath Equalization

At the 900 MHz range, radio waves bounce off everything-buildings, hills, cars, airplanes, etc. Thus many reflected signals, each with a different phase, can reach an antenna. Equalization is used to extract the desired signal from the unwanted reflections. It works by finding out how a known transmitted signal is modified by multipath fading, and constructing an inverse filter to extract the rest of the desired signal. This known signal is the 26-bit training sequence transmitted in the middle of every time-slot burst. The actual implementation of the equalizer is not specified in the GSM specifications.

1.4.5 Frequency Hopping

The mobile station already has to be frequency agile, meaning it can move between a transmit, receive, and monitor time slot within one TDMA frame, which normally are on different frequencies. GSM makes use of this inherent frequency agility to implement slow frequency hopping, where the mobile and BTS transmit each TDMA frame on a different carrier frequency. The frequency hopping algorithm is broadcast on the Broadcast Control Channel. Since multipath fading is dependent on carrier frequency, slow frequency hopping helps alleviate the problem. In addition, co-channel interference is in effect randomized.

1.4.6 Discontinuous Transmission

Minimizing co-channel interference is a goal in any cellular system, since it allows better service for a given cell size, or the use of smaller cells, thus increasing the overall capacity of the system. Discontinuous transmission (DTX) is a method that takes advantage of the fact that a person speaks less that 40 percent of the time in normal conversation, by turning the transmitter off during silence periods. An added benefit of DTX is that power is conserved at the mobile unit.

The most important component of DTX is, of course, Voice Activity Detection. It must distinguish between voice and noise inputs, a task that is not as trivial as it appears, considering background noise. If a voice signal is misinterpreted as noise, the transmitter is turned off and a very annoying effect called clipping is heard at the receiving end. If, on the other hand, noise is misinterpreted as a voice signal too often, the efficiency of DTX is dramatically decreased. Another factor to consider is that when the transmitter is turned off, there is total silence heard at the receiving end, due to the digital nature of GSM. To assure the receiver that the connection is not dead, *comfort noise* is created at the receiving end by trying to match the characteristics of the transmitting end's background noise.

1.4.7 Discontinuous Reception

Another method used to conserve power at the mobile station is discontinuous reception. The paging channel, used by the base station to signal an incoming call, is structured into sub-channels. Each mobile station needs to listen only to its own sub-channel. In the time between successive paging sub-channels, the mobile can go into sleep mode, when almost no power is used.

1.4.8 Power Control

There are five classes of mobile stations defined, according to their peak transmitter power, rated at 20, 8, 5, 2, and 0.8 watts. To minimize co-channel interference and to conserve power, both the mobiles and the Base Transceiver Stations operate at the lowest power level that will maintain an acceptable signal quality. Power levels can be stepped up or down in steps of 2 dB from the peak power for the class down to a minimum of 13 dBm (20 milliwatts).

The mobile station measures the signal strength or signal quality (based on the Bit Error Ratio), and passes the information to the Base Station Controller, which ultimately decides if and when the power level should be changed. Power control should be handled carefully, since there is the possibility of instability.

This arises from having mobiles in co-channel cells alternating increase their power in response to increased co-channel interference caused by the other mobile increasing its power. This in unlikely to occur in practice but it is (or was as of 1991) under study.

2. GSM SECURITY

2.1 Overview

Security was perceived by some in the first days of GSM as an unnecessary expense. Certainly, initially, all involved considered protection of user data from eavesdropping as more important than authentication of the user, through some questioned whether the perceived complexity of introducing encryption over the radio interface was justified. However, as fraud losses from cloning of analogue phones rocketed in the US and the UK, as the Dutch PTT withdrew all its NMT phones so a form of authentication could be added, and as Germans introduced simple authentication on its C-Netz system, it became apparent that authentication of user identity was also very important.

In recent times, the security of GSM has been attacked as too weak. These criticisms are often made without knowledge of neither, the design goals of GSM security nor the regulatory context in which the designers had to work. This chapter aims to show that GSM security met its design goals in a simple and elegant way, and has provided more than adequate security for most of its users. Indeed, GSM offers more "access network" security than fixed phones in most countries (taking the phone to the local exchange link as the "access network" for fixed line systems).GSM has never been subject to commercial cloning that was visited upon analogue NMT, AMPS and TACS system. Moreover, GSM represented the first time ever that encryption functionality had been provided in a consumer device, and played its part in the liberalization of policy on encryption that today's security designers enjoy.

Securing information from unauthorized access is a major problem for any network wire line or wireless. Security, in a broad sense, focuses on network security, system security, information security, and physical security. It is made up of a suite of multiple technologies that solve numerous authentication, information integrity, and identification problems. The mobile telephone has become a necessity for the business professionals or even the general public. Although it provides the convenience of being able to contact people anywhere, a major drawback cannot be ignored - the mobile telephone communication uses radio signal, which is subject to eavesdropping. The mobile telephone network is also subject to other unwanted security breaches, which cost hundreds of million dollars a year in the mobile phone business. There are new technologies emerging which provides a data transfer between different kinds of devices that are close to each other. Security services are needed to offer for mobile users without possibilities to violate the general trust.

2.2 Origins of GSM Security

The security of GSM was developed by the Security Experts Group (SEG) which was formed by CEPT in 1984. There was a lot of concern in CEPT regarding protection of communications systems in general at that time. The origin of the SEG could be said to be a joint meeting of the three CEPT groups CD (data), CS (signaling) and SF (services and facilities) in Berne in January 1984, in which land mobile systems were discussed for the first time. In November, 1984, a proposal from CD to set up a joint CD-GSM group on security (SEG) was accepted by GSM and the first meeting of SEG was held in Malmoe, Sweden, in May 1985.This was a memorable meeting for the delegates as the Swedish air traffic controllers were on strike at that time, forcing the delegates to fly to Copenhagen and travel by boat to Malmoe. The SEG was initially a joint CD/GSM activity, but gradually, the CD part vanished so it was in fact a subgroup of GSM.

The SEG was chaired by Thomas Haug of Swedish Telecommunications A dministration, now called Telia. The membership, like that of CEPT and GSM was drawn from national PTTs and from those organizations that had won a mobile network licence in their country.

2.3 Security Services

When two or more parties exchange information in a session, then authenticity, confidentiality and integrity are needed. Firstly, the parties need to be sure about the identities of each other. Authentication means that the participants somehow prove that their identities are what they claim them to be. Secondly, an outsider should not be able to read the messages. This is meant by confidentiality. Thirdly, integrity requires that messages should not be altered during transmissions. Sometimes also non-repudiation is needed. That is, the sender of a message cannot later deny sending the information, or the receiver cannot deny the reception.

Authentication, integrity, and confidentiality are connected to each other. An authentic message originates from the party it claims to come from, and it arrives unchanged, so this implies integrity. On the other hand, authentication of an intended receiver of a message is used to provide confidentiality. The services in the network may contain classified information, or their computational resources may be limited. Therefore the users may have different access rights. The users need to be authorized to access the services, and access control is needed to prevent unauthorized access.

This way access control provides confidentiality and integrity of the resources. Sometimes the users may need the service to proof the authenticity of the resources as well. A security need for both the sessions and the resources is availability. For the users to be able to communicate with each other and access the resources at all times, high availability is desired. Most security threats fall into the following categories: disclosure of information, loss of integrity, unauthorized access and denial of service.

2.4 Security inside the GSM Network

The GSM security requirements addressed are: data confidentiality over the radio channel, subscriber authentication, and confidentiality of the subscriber's identity and location. The GSM specifications were designed by the GSM Consortium in secrecy and were distributed only on a need-to-know basis to hardware and software manufacturers and to GSM network operators. The specifications were never exposed to the public, thus preventing the open science community around the world from studying the enclosed authentication and enciphering algorithms as well as the whole GSM security model.

The algorithm in question should be publicly available, so that the algorithm is exposed to the scrutiny of the public. According to the general opinion no single entity can employ enough experts to compete with the open scientific community in crypt analyzing an algorithm. Thus, the algorithms designed and implemented in secrecy will probably be somehow cryptographically weak and contain design faults. Eventually, the GSM algorithms leaked out and have been studied extensively ever since by the open scientific community. Interesting facts have been discovered since then, during the cryptanalysis of the A3, A5 and A8 algorithms.

Three distinct security services are provided. These are subscriber identity authentication, user and signalling data confidentiality, and subscriber identity confidentiality. Each of these is considered in turn, and the mechanisms used to provide them outlined. Actually the second of the services is a grouping of three GSM features: user data confidentiality on physical connections, connectionless user data confidentiality, and signalling information element confidentiality. The reason for combining them into one service is that they are all provided by one and the same mechanism.

2.4.1 Subscriber Identity Authentication

This subscriber identity authentication service is the heart of the GSM security system. It is used to enable the fixed network to authenticate the identity of mobile subscribers, and to establish and manage the encryption keys needed to provide the confidentiality services. All networks and mobiles must support the service, although the frequency of application is at the discretion of the network.

Authentication is initiated by the fixed network, and is based upon a simple challenge-response protocol. When a mobile subscriber (MS) attempts to access the system, the network issues it a random challenge RAND. The MS computes a response SRES to RAND using a one-way function A3 under control of a subscriber authentication key K_i. The key K_i is unique to the subscriber, and is shared only by the subscriber and an authentication centre that serves the subscriber's home network. The value SRES computed by the MS is signaled to the network, where it is compared with a pre-computed value. If the two values of SRES agree, the mobile subscriber has been authenticated, and the call is allowed to proceed. If the values are different, then access is denied.

The same mechanism is also used to establish a cipher key K_c for encrypting user and signaling data on the radio path. The key is computed by the MS using a one-way function A8, again under control of the subscriber authentication key K_i, and is pre-computed for the network by the authentication centre that serves the subscriber's home network. Thus at the end of a successful authentication exchange, both parties possess a fresh cipher key K_c.

The pre-computed triples (RAND, SRES, K_c), held by the fixed networks for a particular subscriber, and is passed from the home network's authentication centre to visited networks upon demand. The challenges are used just once. Thus the authentication centre never sends the same triple to two distinct networks, and a network never re-uses a challenge.

In practice the two functions A3 and A8 are combined into a single algorithm, called A38, which is used to simultaneously compute SRES and K_c from RAND and K_i. In this project this combined algorithm is referred to as the authentication algorithm. The protocol described above makes it quite clear that this algorithm need only be available to an authentication centre and the mobile subscribers that that authentication centre serves. In particular, there is no need for a common GSM authentication algorithm and different networks may use different algorithms. (The algorithms do, however, need to have the same input and output parameters; in particular, the length of K_c is determined by the GSM cipher algorithm). Nevertheless it is desirable that there is a GSM standard authentication algorithm, which may be used by all networks, which do not wish to develop a proprietary algorithm. The operators may free to design their A3 algorithm.

2.4.2 User and Signalling Data Confidentiality

As mentioned earlier, this service consists of three elements: user data confidentiality and signaling information on physical connections, connectionless user data confidentiality and signaling information element confidentiality. The first element provides for privacy of all user-generated data, both voice and non-voice, transferred over the radio path on traffic channels. The second element provides for privacy of user data transferred in packet mode over the radio path on a dedicated signaling channel, whilst the third element provides for privacy of certain user related signaling elements transferred over the radio path on dedicated signaling channels.

All of these elements of service are provided using the same layer 1 encryption mechanism, and must be supported and used by all networks and mobiles. The mechanism is now briefly described. Encryption is achieved by means of a ciphering algorithm A5 that produces a key stream under control of a cipher key K_c. This key stream is then bit-for-bit exclusive-or and with the data transferred over the radio path between the MS and the base station (BS). The cipher key is established at the MS as part of the authentication procedure, as described in the last section, and is transferred through the fixed network to the BS after the MS has been identified.

It is essential that the MS and BS synchronize the starting of their cipher algorithms, but this only directly addresses the situation when the network initiates an authentication check. The procedures still need to be specified in detail to cover the situation when the network does not authenticate the MS. When the network intends to issue an authentication challenge, the BS starts deciphering all data immediately after the MS has been identified using the cipher key K_c that the MS will derive upon receipt of the challenge RAND. The MS starts ciphering and deciphering the moment it has computed K_c (and SRES) from RAND, as described in the last section, and before SRES is transmitted. On the BS side, enciphering starts as soon as SRES has been received, deciphered and found to be correct. To cope with possible transmission loss or errors, the authentication request and response message are repeated under the control of timers.

Synchronization of the ciphering key stream is maintained by using the TDMA frame structure of the radio sub-system. The TDMA frame number is used as a message key for the cipher algorithm AS, and the algorithm produces a synchronized key stream for enciphering and deciphering the data bits in the frame. For each frame, a total of 114 bits are produced for enciphering / deciphering data transferred from the MS to the BS, and an additional 114 bits are produced for deciphering / enciphering data received at the MS from the BS. A frame lasts for 4.6 ms, so that the cipher has to produce the 228 bits in this time. The cipher algorithm A5 must be common to all GSM networks, and three algorithms have been proposed as candidates for the GSM standard: a French algorithm, a Swedish algorithm and a UK algorithm.

2.4.3 Subscriber Identity Confidentiality

Subscriber identity confidentiality on the radio interface was one of the security requirements of GSM.

This service allows mobile subscribers to originate calls, update their location, etc, without revealing their International Mobile Subscriber Identity (IMSI) to an eavesdropper on the radio path. It thus prevents location tracing of individual mobile subscribers by listening to the signaling exchanges on the radio path. All mobiles and networks must be capable of supporting the service, but its use is not mandatory.

A robust identity confidentiality mechanism is in fact quite a difficult thing to achieve. Confidentiality usually involves encryption and encryption requires, for symmetric ciphering, a shared secret key. However, generation of a shared secret key should generally be done in combination with, or after authentication, or how does one entity know who they are sharing a secret key with and who therefore they are revealing their identity too? Authentication, however, is authentication of a particular identity, but identities have not yet been exchanged!

Mechanisms involving public key cryptography can be used but only if one of the entities (e.g. a "server" or a network operator) does not mind revealing its identity to eavesdroppers.

A simple mechanism involving public keys might be that one entity (the "server") transmits a certificate for its public key and the other entity encrypts its identity using the received public key. The transmitted identity can then be authenticated by a variety of means that do not reveal the identity to passive eavesdroppers.

Public key cryptography was not available to the GSM desirers, so a simple mechanism, using temporary identities and the basic facilities of GSM security was designed.

In order to provide the subscriber identity confidentiality service it is necessary to ensure that the IMSI, or any information which allows an eavesdropper to derive the IMSI, it not (normally) transmitted in clear in any signaling message on the radio path. The mechanism used to provide this service is based on the use of a temporary mobile subscriber identity (TMSI), which is securely updated after each successful access to the system. Thus, in principle, the IMSI need only be transmitted in clear over the radio path at registration. When a subscribers attempts access with an operator with which it is not presently registered (so, first access in a roamed to network, or the first access for time in its home network) it must reveal its identity, and request access using its permanent identity, the International Mobile Subscriber Identity, or IMSI. The IMSI is then authenticated, a process which results in the sharing of K_c. The subscriber is then assigned a Temporary Mobile Subscriber Identity (TMSI, pronounced "timsy") which is sent to the subscriber encrypted with K_c. The next time the user attempts access in that network, it uses the TMSI to identify itself and the network looks up its table of TMSI to IMSI mapping to find the subscriber's permanent identity and the triplets with which it can authenticate the subscriber and begin encryption. So that a subscriber cannot be followed around, it is frequently given a new TMSI (if the same TMSI were used for a while, a subscriber previously identified by some out of band means could be recognized by the TMSI). Theoretically, the IMSI should only have to be used on a subscriber's first ever registration with any network, and it should be possible for the TMSI to be used even across different networks. In practice, however, the IMSI must be revealed on first registration in a new network at least, and in some networks, more frequently than this.

The GSM identity confidentiality is simple and efficient, but is not robust. The IMSI must be revealed on first registration with a network, and the mechanism as a whole can be compromised using a " false base station".

The TMSI updating mechanism functions in the following manner. For simplicity, assume the MS has been allocated a TMSI, denoted by TMSIo, and the network knows the association between TMSIo and the subscriber's IMSI. The MS identifies itself to the network by sending TMSIo. Immediately after authentication (if this takes place), the network generates a new TMSI, denoted TMSIn, and sends this to the MS encrypted under the cipher key k_c. Upon receipt of the message, the MS deciphers and replaces TMSIo by TMSIn.

2.5 GSM Security Model

2.5.1 The Purpose of GSM Security:

The use of radio communications for transmission to the mobile subscribers makes GSM Public Land Mobile Networks (PLMN) particularly sensitive to misuse of their resources by unauthorized persons using manipulated Mobile Stations, who try to impersonate authorized subscribers and eavesdropping of the various information, which are exchanged on the radio path.

So the security features in GSM PLMN is implemented to protect:

- The access to the mobile services.
- Any relevant item from being disclosed at the radio path, mainly in order to ensure the privacy of user-related information.

Since all cases of GSM fraud against a specific wireless carrier will result in a substantial loss to the operator. This substantial loss may include the following:

- No direct financial loss, where the result is lost customers and increase in use of the system with no revenue.
- Direct financial loss, where money is paid out to others, such as other networks, carriers and operators of 'Value Added Networks' such as Premium Rate service lines.
- Potential embarrassment, where customers may move to another service because of the lack of security.
- Failure to meet legal and regulatory requirements, such as License conditions, Companies Acts or Data Protection Legislation

2.5.2 GSM's Security Limitations

Existing cellular systems have a number of potential weaknesses that were considered in the security requirements for GSM.

The security for GSM has to be appropriate for the system operator and customer:

- The operators of the system wish to ensure that they could issue bills to the right people, and that the services cannot be compromised.
- The customer requires some privacy against traffic being overheard.

The countermeasures are designed to:

- make the radio path as secure as the fixed network, which implies anonymity and confidentiality to protect against eavesdropping;
- have strong authentication, to protect the operator against billing fraud;
- prevent operators from compromising each others' security, whether inadvertently or because of competitive pressures.

The security processes must not:

- significantly add to the delay of the initial call set up or subsequent communication;
- increase the bandwidth of the channel,
- allow for increased error rates, or error propagation;
- add excessive complexity to the rest of the system,
- must be cost effective.

The designs of an operator's GSM system should take into account, the environment and have secure procedures such as:

- the generation and distribution of keys,
- exchange of information between operators,
- the confidentiality of the algorithms.

2.5.3 Descriptions of the functions of the services

The security services provided by GSM are:

- **Anonymity** so that it is not easy to identify the user of the system.
- **Authentication** so the operator knows who is using the system for billing purposes.
- **Signalling Protection** so that sensitive information on the signaling channel, such as telephone numbers, is protected over the radio path.
- **User Data Protection** so that user data passing over the radio path is protected.

Anonymity

Anonymity is provided by using temporary identifiers. When a user first switches on his/her radio set, the real identity is used, and a temporary identifier is then issued. From then on the temporary identifier is used. Only by tracking the user is it possible to determine the temporary identity being used.

Authentication

Authentication is used to identify the user (or holder of a Smart Card) to the network operator. It uses a technique that can be described as a "Challenge and Response", based on encryption.

Authentication is performed by a challenge and response mechanism. A random challenge is issued to the mobile, the mobile encrypts the challenge using the authentication algorithm (A3) and the key assigned to the mobile, and sends a response back. The operator can check that, given the key of the mobile, the response to the challenge is correct.

Eavesdropping the radio channel reveals no useful information, as the next time a new random challenge will be used. Authentication can be provided using this process. A random number is generated by the network and sent to the mobile. The mobile use the Random number R as the input (Plaintext) to the encryption, and, using a secret key unique to the mobile K_i, transforms this into a response Signed RES ponse (SRES) (Cipher text) which is sent back to the network.

The network can check that the mobile really has the secret key by performing the same SRES process and comparing the responses with what it receives from the mobile.

User Data and Signalling Protection

The response is then passed through an algorithm A8 by both the mobile and the network to derive the key K_c used for encrypting the signalling and messages to provide privacy (A5 series algorithms).

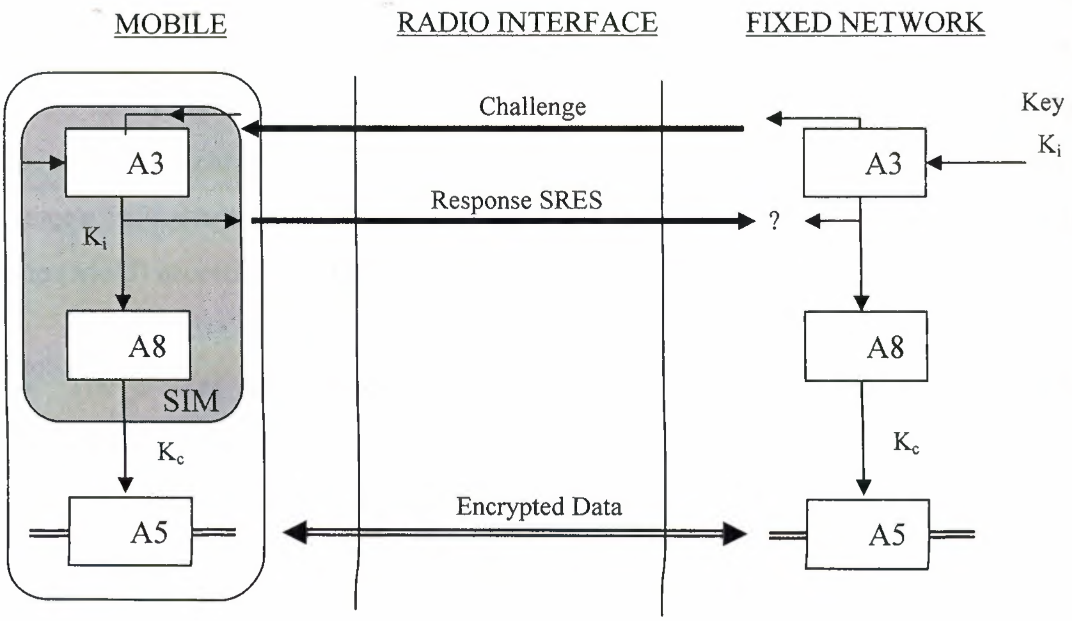

Figure 2.1 Encryption for GSM

Implementation and Roaming

The authentication algorithm A3 is an operator option, and is implemented within the smart card (known as the Subscriber Interface Module or SIM). So that the operators may inter-work without revealing the authentication algorithms and mobile keys (K_i) to each other, GSM allows triplets of challenges (R), responses (SRES) and communication keys (K_c) to be sent between operators over the connecting networks.

The A5 series algorithms are contained within the mobile equipment, as they have to be sufficiently fast and are therefore hardware. There are two defined algorithms used in GSM known as A5/1 and A5/2. The enhanced Phase 1 specifications developed by ETSI allows for inter-working between mobiles containing A5/1, A5/2 and unencrypted networks. These algorithms can all be built using a few thousand transistors, and usually takes a small area of a chip within the mobile.

2.5.4 World-Wide Use of the Algorithms

There are now three different possibilities for GSM, unencrypted, and use of the A5/1 algorithm or the A5/2 algorithm to secure the data. This arose because the GSM standard was designed for Western Europe, and export regulations did not allow the use of the original technology outside Europe. The uses of the algorithms in the network operator's infrastructure are controlled by the GSM Memorandum of Understanding Group (MoU) according to the formula below:

- The present A5/1 algorithm can be used by countries which are members of CEPT.
- The algorithm A5/2 is intended for any operators in countries that do not fall into the above category.

Export controls on mobiles are minimal, and the next generation of mobiles will support A5/1, A5/2 and no encryption. The protocols to support the various forms of A5 (up to seven) are available in GSM.

Loss areas

There are a number of areas that can be exploited; the most likely intention of all the techniques is the ability to make money at the lowest cost possible.

Technical fraud

Technical fraud is where a weakness of the system is exploited to make free calls. For example, Call Forwarding or Conference Call facilities may be used to give reduced price services to customers from a stolen mobile. These are often known as 'Call Sales Offices'. Hackers and phreakers are often able to gain access and exploit a weakness in the switching or billing system and gain the ability to make calls or financial advantage. In some cases hackers and phreakers can take over the entire billing system and routing system; thus causing convenience for customers and carriers.

Procedural fraud

Procedural fraud results from the exploitation of business processes, where a flaw or weakness can be used to gain money. It may be possible for example to get free calls from a stolen mobile, and sell the calls on for a lower cost than any legitimate network operator. This can be minimized by designing processes so that losses can be stopped by the use of correct and up to date policies, and by taking the opportunity to create a fraud away from the attacker or employee.

Comparison with other frauds

Many of the techniques that can be used to commit fraud on telecommunications networks can also be used for a mobile network. Analogue mobile phone systems (AMPS) were subject to being eavesdropped (with conventional RF-Scanners available at electronics shops and Radio Shack), and the phones could be cloned (ESN snarling over thin-air) so that bills were paid by the owner of the original mobile phone. Existing cellular systems have a number of potential weaknesses that were considered in the security requirements for GSM. Networks such as GSM, with international roaming and interactions with other operators (carriers), offer other opportunities for exploitation. GSM has been designed to offer various technical solutions to prevent misuse, such as strong authentication, together with anonymity and encryption of the signaling and data over the radio. However, all systems are dependent on secure management deployment and special procedures; lapses in these areas have severe impact on the resilience of the business process to fraud. For example; many carriers still make use of the COMP128 encryption algorithm for both A3 (the authentication algorithm to prevent phone cloning) and A8 (the voice-privacy key-generation algorithm), which is fine for securing against simple over-the-air attacks.

However we have determined that the COMP128's voice-encryption algorithms only encrypt voice between the GSM wireless phone and the base station. It does not encrypt voice within the phone network, nor does it encrypt end to end. It only encrypts the over-the-air portion of the transmission. The attack on COMP128 takes just 2^19 queries to the GSM smart-card chip, which takes approximately 8 hours over the air. This attack can be tested on as many simultaneous phones in radio range as your rogue base station has channels.

2.6 Design Goals

The security functionality within any system is a balance between the likelihood and impact of threats, user demand for certain security features and the cost and complexity of security measures. A security mechanism that is impervious to attack by any organization over any timescale would generally not be appropriate goals for a system transporting public, largely non-sensitive data. System designers must therefore set appropriate goals for the security of their system prior to beginning detailed design. SEG undertook this task and came up with the following simple goal for GSM security:

It would provide a degree of protection on the radio path which was approximately the same as that provided in the fixed network.

SEG were concerned with security on the radio interface only – there was no attempt to provide security on the fixed network part of GSM.

Before describing how this simple goal was translated into more formal security requirements, a few definitions are given:

Confidentiality is the property data has when it cannot be read by parties not authorized to read it. Confidentiality is provided by encryption in GSM.

Authentication of user identity is the property of establishing that the claimed identity of an entity really is their identity.

Integrity protection is the property of data whereby modification to the data can be detected. This is not explicitly provided by GSM but is provided implicitly by the use of ciphering along with the use of non-linear checksums (as stream ciphers are used in GSM, stream ciphering along does not provide integrity protection).

The following requirements for the GSM security were developed over the course of the design exercise. These are listed in:

- Subscriber identity authentication. This protects the network from unauthorized use.
- Subscriber identity confidentiality. This provides protection against the tracing of a user's location by listening to exchanges on the radio interface.

- User data confidentiality across the radio interface. This protects the user's connection orientated data from eavesdropping on the radio interface.
- Connectionless user data confidentiality across the radio interface. This protects user information sent in connectionless packet mode in a signaling channel from eavesdropping on the radio interface.
- Signaling information element confidentiality across the radio interface. This protects selected fields in signaling messages from eavesdropping o the radio interface.

There was not general agreement on the issue of identity confidentiality within the group. Some members felt it was very important, particularly the German delegates. Other felt it was not a real requirement, and the subscriber must in some circumstances reveal their identity anyway, that the requirement could not be robustly met in any case.

There was also some debate during the design of GSM security as to whether user data should be given " privacy" or "confidentiality". "Privacy" was taken to mean protection from a determined "amateur" attacker but not necessarily a large organization – "confidentiality" was taken to mean protection from attack by the latter. The final conclusion was to try and provide confidentiality.

It should be noted that right from the start, there was concern within the group of providing too much security and thereby bringing unnecessary export problems upon GSM. The security was therefore designed with this constraint in mind, and also two further constraints:

- GSM did not have to be resistant to "active attacks" where the attacker interferes with the operation of the system, perhaps masquerading as a system entity. Active attacks are in contrast to passive attacks where the attacker merely monitors inter-system communication and does not interfere.
- The trust that must exist between operators for the operation of the security should minimized.

2.7 Choosing the Security Aarchitecture for GSM

Many people see security for communications systems as a matter of algorithms and attacks on the security of communications systems as a matter of attacks on algorithms. However, in designing security for a system, the choice of algorithms is often one of the last choices. The first task is to decide the goals of the security, which for GSM we have already talking about.

The second choice is determine the security protocols that will be used to achieve these goals. Usually, only after this point can the algorithms be decided. However, if the choice of algorithm is going to involve the choice between the use of secret or public key cryptography, then this basic choice must be made sooner, as it may dictate the whole security architecture.

The choice of public or secret key cryptography should still not be taken until the security goals have been decided though, in all the interesting debates about algorithm security, the designers may lose sight of the goals and why they are actually engaged in a design process at all. Having said all this, there were contributions at SEG meetings which proposed particular architectures without any rationale or reference to claimed goals.

A security protocol is an interaction between two or more (but usually only two) entities following pre-determined steps that achieve some security goal. These protocols may involve proving that certain parties have certain items of secret information (this occurs during authentication or proof of claimed identity) and also may involve distribution or generation of secret keys for protection of communication. For instance, the widely known protocol, SSL achieves authentication of the server (generally a web server), generation of a shared secret key of protect the communication of data between the client (usually a browser on a PC) and the server, and the subsequent protection of that data for the duration of the SSL session. The emphasis on server authentication (though client authentication is possible) in SSL was provided so that users would have confidence in who they were sending data, e.g. credit card numbers, to. Client authentication is not mandatory in SSL as the use of the channel is generally "free" or the user has already been authenticated for charging purposes prior to the start of the SSL session (and as authentication is public key based in SSL, there is the complexity of provisioning clients with key pairs and certificates).This emphasis on server authentication can be contrasted with the emphasis on client or user authentication in GSM, where the use of the channel is not free, and the user must therefore be authenticated so that they can be charged.

There are many ways of satisfying the goals for GSM security and the process of designing the GSM security architecture reflected the many possibilities open to the designers – many candidate architectures were proposed by the participating parties. BT, for instance, proposed the use of public key cryptography along with their own secret, symmetric encryption algorithm, BeCrypt. The reader might be interested to know why public Key Cryptography (PKC) was not used. There were three main reasons:

- Implementations at the time were immature, the impact on the terminal for the provision of PKC functionality was therefore not accurately know:
- Messages would be longer, as PKC requires longer keys than symmetric cryptography, for provision of the same cryptographic strength;
- There was no real gain from the use of PKC. The authentication protocol runs between a subscriber and the network operator the subscriber has chosen to use. There is therefore a well established relationship and the one to many authentication possibility of PKC is not therefore required.

The large number of proposals caused problems for the group's progress, in that there were just too many protocols to examine properly. A security protocol should not be accepted until it has been examined thoroughly by a good number of experts. Flaws in communications systems security often occur because of a weakness in the protocols involved, and not in the strength of algorithms. However, these flaws are often subtle and take time and careful analysis to uncover. A small group therefore decided to trim the number of proposals down to a manageable level. This small group was, as is often the case in such situations, not elected or formally tasked with trimming the number; it was just a small number of people taking an initiative.

2.8 The Architecture Chosen

The architecture finally chosen was a simple and elegant one. It is based on secret key and not public key cryptography as stated above.

The architecture is centred and a long-term secret key, K_i, which is possessed by both the subscriber's mobile phone and subscriber's operator only. Authentication of the mobile phone by the network consists of proof by the mobile phone that it possesses the K_i. As part this process cipher keys used for encryption during a call are also derived from k_i.

Before describing these operations in detail, a couple of design principles /constraints must be given. The first is that it was decided that the k_i must remain with the subscriber's home operator and most not be passed to other operator if the subscriber's home operator and must not be passed to anther operator if the subscriber roams to that network this is because the k_i is such a sensitive piece of information.With the k_i of a particular subscriber an impostor can pretend to be ("masquerade as ") that subscriber and they can eavesdrop on all subscriber's calls k should therefore not be revealed to more entities than is strictly necessary and SEG found a way for it only to be known by the minimum number of entities , the mobile phone (actually the "SIM", see below) and the home network. The second constraint is that long distance signalling should be minimized. It was therefore not acceptable that authentication process should involve the home operator for every call made by a roaming subscriber of that operator.

K_i should not every be known by the user themselves either, as this would allow the self-cloning of phone, and subsequent denial of the calls made by the cloned phone that had already occurred in analogue networks. A secure module within the phone that could be programmed by or under the control of the operator and in which k was stored and all operation involving k carried out was therefore required. A smart card was the obvious choice for such a security module, and the GSM Subscriber Identity Module (SIM) was born .The k_i is stored and used in the SIM and not in the terminal (or to use GSM terminology, the mobile equipment (ME)).

The security architecture is now described, in stages.

K_i in the home operator is held in the operator Authentication Center (AuC).(GSM phase 1did allow the k_i to be sent from the AuC to VLR for use there, but this way only allowed for VLRs in the same PLMN as the AuC, the specification advised against the option, and the option was dropped in GSM Phase2.) The AuC generates a random number, RAND for each subscriber. Random challenges are commonly used in security protocols to guarantee that a particular run of the protocol is "fresh" and entirely new and that an impostor who has captured some parameters from a previous run of the protocol cannot masquerade as the genuine subscriber or operator or interfere(either actively or passively) with the current run of the protocol. As shown in figure2.2, for a particular subscriber,each RAND is passed as a parameter, a long with the k_i for that subscriber, through an algorithm named A3. A3 produces as an output, an expected response, XRES. The use of a challenge-response mechanism was not a proposal of a particular delegate in SEG. Once it was decided that a secret key mechanism would be, a challenge-response mechanism was the obvious choice.

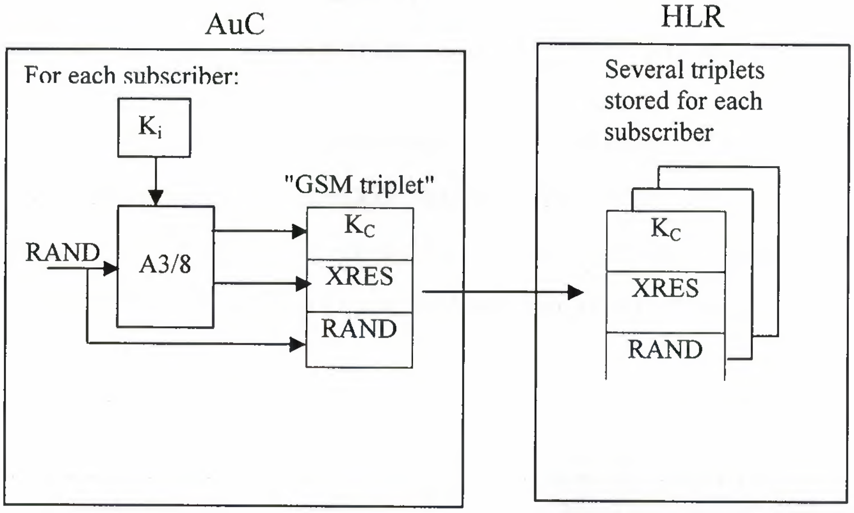

Figure 2.2 Diagram of Triplet Generation

RAND and k_i are also passed to another algorithm A8 which produces a cipher key, K_c. Typically, algorithms A3 and A8 are combined into one, called A3/8, and we shall consider them as such from now on. A RAND and the resulting XRES and Kc produced by A3/8 are called a "triplets". An AuC will normally produce a batch of triplets for a particular subscriber all at once and pass these for distribution to the HLR. This separation of triplet generation in the AuC from triplet distribution and subscriber management in the HLR, means that the AuC need only communicate with the HLR. In theory, therefore, greater access control can be placed on the AuC since it only ever communicates with one, known, entity, the associated HLR of the same operator. When a subscriber attempts to make a call or a location update in either its home operator's network, or in a network it has roamed to, the SIM passes its identity to the VLR serving that subscriber. The VLR makes a request to the subscriber's HLR for a batch of triplets for the identity claimed by the subscriber (i.e. the SIM) and the HLR responds with a batch of triplets for that claimed identity. The VLR authenticates the SIM by sending a RAND from the batch to the mobile phone, as shown in Figure 2.2. The ME passes RAND to the SIM where K_i and A3/8 are held. The SIM passes RAND and its K_i through algorithm(s) A3/8 residing within the SIM as was done in the AuC. The "signed response" produced by the SIM, SRES, is passed back to the VLR. The VLR compares SRES with the expected response, XRES, for that RAND, and if they match, the SIM/mobile phone's claimed identity is deemed to be authenticated. A3/8 in the SIM also produces K_c and if the SIM is authenticated, the VLR passes the K_c from the triplet to the Base Transceiver Station (BTS, the "abase station") serving the mobile. The SIM passes K_c to the ME and the BTS and mobile can then begin ciphering communications using K_c. The algorithm used for ciphering is termed A5.

With the use of the triplets, authentication can be performed in the serving network without the serving network operator having knowledge of K_i. When the serving network has run out of triplets, it should request more from the home operator (though the serving network is allowed to re-use triplets if it cannot obtain more).

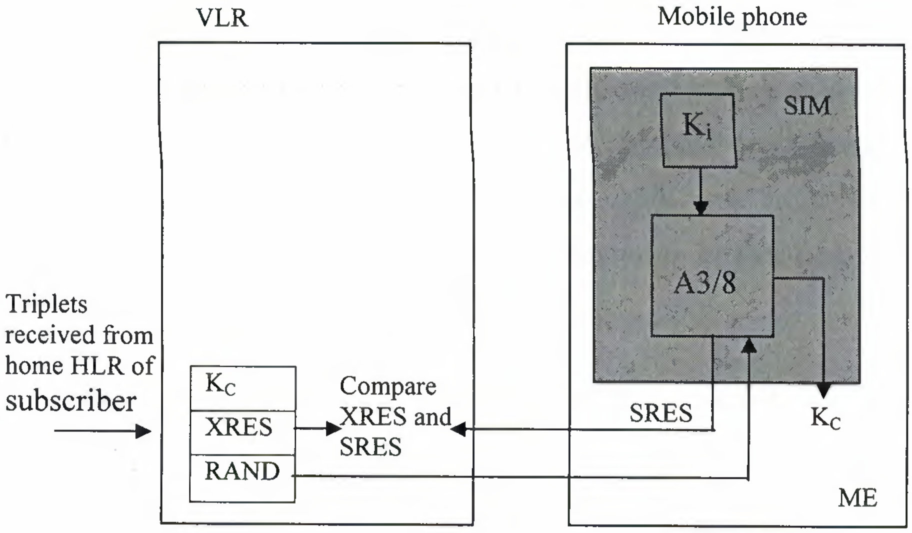

Figure 2.3 Diagram for authentication in the serving VLR

The inquisitive reader may be wondering when the use of algorithms A1-4, A6 and A7 is to be described. Their us, will not be described though, as they, and also algorithms A9 to A12 came up in initial versions of the architecture, but were not required in the final version. A3, A5 and A8 were not renamed A1 to A3. A4 and K4 are used by some operators to denote the encryption algorithm and key protecting the personalization data of a SIM (including the IMSI and K_i) between the personalization centre and the AuC, but they are not specified in any GSM specification.

The security architecture described above was specified in GSM specification 03.20

2.9 Authentication Algorithm Design

The effectiveness of authentication relies on a number of algorithm requirements not yet given. The first is that it is statistically near impossible for an impostor to guess what the correct SRES should be and therefore masquerade as another subscriber. As parameters SRES/XRES are 32 bits long, and the mobile has only one chance to return SRES for a particular RAND, provided that the algorithm has been so designed that SRES is indistinguishable from any other 32 bit number that might be returned instead of SRES, such an impostor has only a 1 in 2^{32}, or 1 in approximately 10 billion chance of guessing SRES correctly. This was felt sufficiently improbable as to not represent a realistic attack.

The second assumption is that, as RAND and SRES are passed un-encrypted between the mobile and the base station, an impostor cannot derive K_i from collecting a number of RAND-SRES pairs. This means that A3/8 must be designed to resist a known plaintext attack where the attacker knows what is ciphered as well as the ciphered result. Further, as an attacker could steal a SIM for some time, and send whatever challenges he liked to the SIM, and collect the SRESs given, A3/8 must be resistant to a chosen plaintext attack. This latter requirement was shown not to be satisfied by the algorithm COMP128, used as A3/8 by many operators.

A third requirement is that, again as RAND and SRES are passed un-encrypted between the mobile and the base station, an impostor cannot derive a particular K_c from the RAND and SRES in the same triplet as that K_c or by collecting a number of RAND-SRES pairs. This means that SRES and K_c must be completely unrelated though derived from the same RAND and K_i.

It has been mentioned that an important design consideration was that K_i was not to be shared with the serving network. A by-product of this decision is that algorithm A3/8 does not need to be known by the serving VLR, as A3/8 is only used where K_i is present, that is, in the AuC and the SIM. It should be noted that the VLR does not need any cryptographic algorithms, as A5 is not used in the VLR either but in the BTS, the security functionality in the VLR is therefore a dimple comparison and distribution of parameters. This dicers from systems based on ANSI-41, as used in many US networks, where the VLR must possess cryptographic capability. As 43/8 is only present in tie SIM and AuC and the use of A3/8 is a protocol between a subscriber's SIM and the AUC of that subscriber's operator (albeit with the HLR and VLR as intermediaries), A3/8 does not have to be standardized. However, as the parameters of the triplet are passed via the HLR, VLR and ME as well as the SIM and AuC, the lengths of the parameters in the triplets must be standardized in the absence of a flexible encoding method. Each operator can therefore have a different A3/8 and operators were encouraged to take advantage of this possibility.

SEG felt it was an advantage that A3/8 did not need to be standardized. One claimed advantage of this is that less standardization work must be done. However, in response to this it could be said that now each operator must develop their own A3/8, so thought the amount of standardization has gone down, the amount of development will go up. A second purported advantage is that each operator can also keep their A3/8 secret. However, this is also a moot point, because, as has been mentioned previously with regard to protocols, flaws in algorithms can be very subtle, and keeping an algorithm secret necessarily means there will be less potential examination of the algorithm. A clear advantage is that operators can gracefully bring in a new A3/8 on a SIM by SIM basis - the AuC knows which subscriber a request for triplets is for and can therefore use an updated A3/8. A clear disadvantage of there being different A3/8 is that AuC: manufacturers must cope with different requirements from different operators.

However, in spite of the arguments for and against standardization of A3/8, it was recognized that an example algorithm would be required, for implementation tests, and for those operators that did not possess or wish to possess the capability to obtain such an algorithm.

This algorithm was COMP128, designed by a research wing of Deutches Telecom. The use of COMP128 amply illustrates the disadvantages mentioned above.

2.10 Other GSM Security Mechanisms

2.10.1 SIM Card

There is always the possibility (we have no knowledge) that the SIM card can be compromised. This is considered unlikely, especially as some operators use their own version of A3. Keys K_i and the matching IMSI could be compromised by someone selling the information for money.

2.10.2 IMEI

In GSM the customer subscription and authentication capability is contained within a smart card (SIM, Subscriber Identity Module). Any mobile will take on the identity of a subscriber by insertion of a smart card. The mobiles now become attractive items to steal, as they can be used with another SIM card.

To prevent this, GSM has specified an International Mobile Equipment Identifier (IMEI). Although to an operator, at first evaluation, it may seem as the stolen mobiles have no effect, as they do not affect a subscription, there will be problems with an increase in customer facing staff to handle esquires, and a possibility that GSM handsets are expensive to insure.

An Equipment Identity Register (EIR) exists in each network, with Black, White and Grey Lists for stolen or non type approved mobiles, valid mobiles and mobiles that need tracking respectively. Grey lists are for local tracking of mobiles within a network.

GSM has defined a procedure so that approved, lost or stolen mobile IMEIs can be communicated to all other operators. A Central Equipment Identity Register has been (CEIR) proposed. Type approval authorities' issue white list numbers (random ranges of valid IMEIs) to mobile manufacturers and manufacturers inform the CEIR when the mobiles are released to market. All operators are able to post their black lists to the CEIR, and in return collect a consolidated list of all operators black and white lists.
By this method stolen or invalid mobiles can be quickly barred throughout the world.

2.10.3 Roaming

International roaming problems are minimised by the use of two procedures:

- Rapid exchange of billing information by means of EDI
- Notification of the home network of the visitor when the visitor has exceeded a certain billing limit.

2.11 Problems with GSM Security

2.11.1 The Limitation and Problems with GSM Security

Problems with GSM security

- Security by obscurity, which means that all of the algorithms used are not available to the public. Most security analysts believe any system that is not subject to the scrutiny of the world's best minds can't be as secure.
- Only provides access security. All communication between the Mobile Station and the Base Transceiver Station are encrypted. But all communications and signalling is generally transmitted in plain text in the fixed network.
- Difficult to upgrade the cryptographic mechanisms
- Lack of user visibility (e.g. doesn't know if encrypted or not)
- The flaw of the algorithms.

2.11.2 Possible Improvement

Security could be improved in some areas with relatively simple measures. One solution is to use another cryptographically secure algorithm for A3. This would require issuing new SIM-cards to all subscribers and updating HLR software. This would effectively disable the attacker from cloning SIM-cards, the most dangerous attack, which is discussed above. This solution is easy to be implemented because the network operators can make the changes themselves and do not need the support of hardware or software manufacturers or the GSM Consortium. There is now a new algorithms available called COMP128-2.

The operator can employ a new A5 implementation with strong encryption too. A new A5/3 algorithm has also been agreed upon to replace the aging A5/2 algorithm. This improvement would require the co-operation of the hardware and software manufacturers because they will have to release new versions of their software and hardware that would comprise with the new algorithm.

Third solution would be to encrypt the traffic on the operator's backbone network between the network components. This would disable the attacker from wiretapping the backbone network.

This solution could probably also be implemented without the blessings of the GSM Consortium, but the co-operation of the hardware manufacturers would still be required.

3. CELLULAR COMMUNICATIONS

3.1 Overview

A cellular mobile communications system uses a large number of low-power wireless transmitters to create cells, the basic geographic service area of a wireless communications system. Variable power levels allow cells to be sized according to the subscriber density and demand within a particular region. As mobile users travel from cell to cell, their conversations are handed off between cells to maintain seamless service. Channels (frequencies) used in one cell can be reused in another cell some distance away. Cells can be added to accommodate growth, creating new cells in un-served areas or overlaying cells in existing areas.

This chapter discusses the basics of radio telephony systems, including both analog and digital systems. Upon completion of this chapter, you should be able to describe the basic components of a cellular system and identify digital wireless technologies.

3.2 Mobile Communications Principles

Each mobile uses a separate, temporary radio channel to talk to the cell site. The cell site talks to many mobiles at once, using one channel per mobile. Channels use a pair of frequencies for communication—one frequency (the forward link) for transmitting from the cell site and one frequency (the reverse link) for the cell site to receive calls from the users. Radio energy dissipates over distance, so mobiles must stay near the base station to maintain communications. The basic structure of mobile networks includes telephone systems and radio services. Where mobile radio service operates in a closed network and has no access to the telephone system, mobile telephone service allows interconnection to the telephone network.

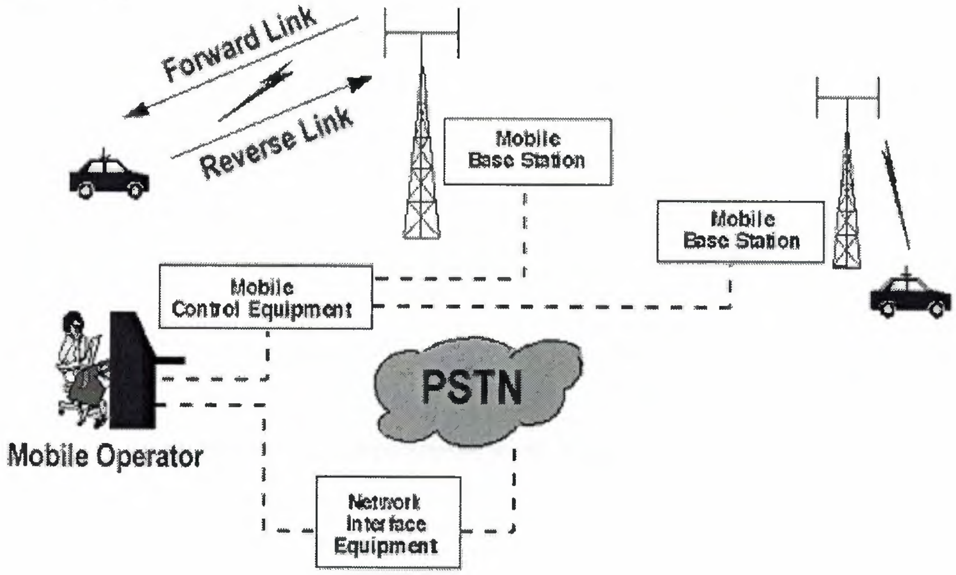

Figure 3.1 Basic Mobile Telephone Service Network

3.2.1 Early Mobile Telephone System Architecture

Traditional mobile service was structured in a fashion similar to television broadcasting: One very powerful transmitter located at the highest spot in an area would broadcast in a radius of up to 50 kilometers. The cellular concept structured the mobile telephone network in a different way. Instead of using one powerful transmitter, many low-power transmitters were placed throughout a coverage area. For example, by dividing a metropolitan region into one hundred different areas (cells) with low-power transmitters using 12 conversations (channels) each, the system capacity theoretically could be increased from 12 conversations or voice channels using one powerful transmitter to 1,200 conversations (channels) using one hundred low-power transmitters. Shows a metropolitan area configured as a traditional mobile telephone network with one high-power transmitter.

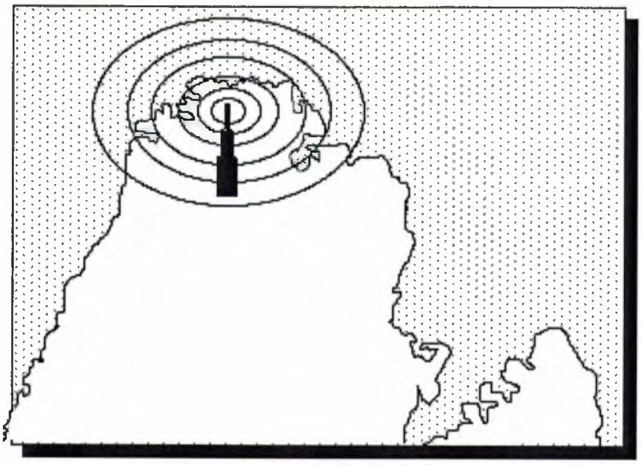

Figure 3.2. Early Mobile Telephone System Architecture

3.3 Mobile Telephone System Using the Cellular Concept

Interference problems caused by mobile units using the same channel in adjacent areas proved that all channels could not be reused in every cell. Areas had to be skipped before the same channel could be reused. Even though this affected the efficiency of the original concept, frequency reuse was still a viable solution to the problems of mobile telephony systems.

Engineers discovered that the interference effects were not due to the distance between areas, but to the ratio of the distance between areas to the transmitter power (radius) of the areas. By reducing the radius of an area by 50 percent, service providers could increase the number of potential customers in an area fourfold. Systems based on areas with a one-kilometer radius would have one hundred times more channels than systems with areas 10 kilometers in radius. Speculation led to the conclusion that by reducing the radius of areas to a few hundred meters, millions of calls could be served.

The cellular concept employs variable low-power levels, which allow cells to be sized according to the subscriber density and demand of a given area. As the population grows, cells can be added to accommodate that growth.

Frequencies used in one cell cluster can be reused in other cells. Conversations can be handed off from cell to cell to maintain constant phone service as the user moves between cells.

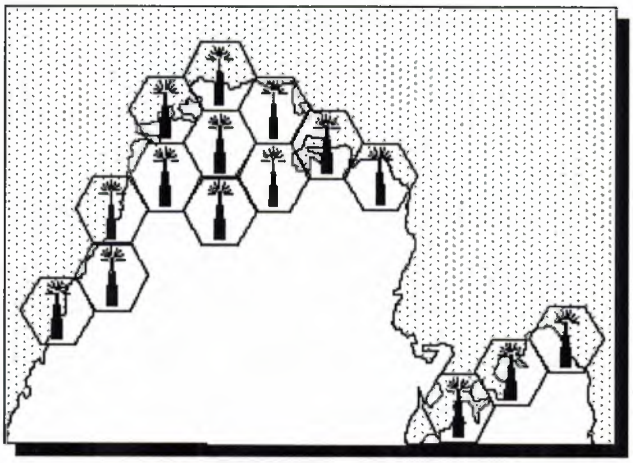

Figure 3.3. Mobile Telephone System Using a Cellular Architecture

The cellular radio equipment (base station) can communicate with mobiles as long as they are within range. Radio energy dissipates over distance, so the mobiles must be within the operating range of the base station. Like the early mobile radio system, the base station communicates with mobiles via a channel. The channel is made of two frequencies, one for transmitting to the base station and one to receive information from the base station.

3.4 Cellular System Architecture

Increases in demand and the poor quality of existing service led mobile service providers to research ways to improve the quality of service and to support more users in their systems. Because the amount of frequency spectrum available for mobile cellular use was limited, efficient use of the required frequencies was needed for mobile cellular coverage.

In modern cellular telephony, rural and urban regions are divided into areas according to specific provisioning guidelines. Deployment parameters, such as amount of cell-splitting and cell sizes, are determined by engineers experienced in cellular system architecture. Provisioning for each region is planned according to an engineering plan that includes cells, clusters, frequency reuse, and handovers.

3.4.1 Cells

A cell is the basic geographic unit of a cellular system. The term *cellular* comes from the honeycomb shape of the areas into which a coverage region is divided. Cells are base stations transmitting over small geographic areas that are represented as hexagons. Each cell size varies depending on the landscape. Because of constraints imposed by natural terrain and man-made structures, the true shape of cells is not a perfect hexagon.

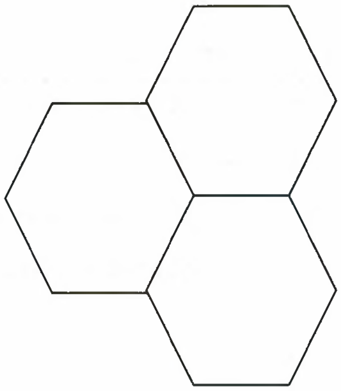

Figure 3.4. Cells

3.4.2 Clusters

A cluster is a group of cells. No channels are reused within a cluster. Figure 3.5 illustrates a seven-cell cluster.

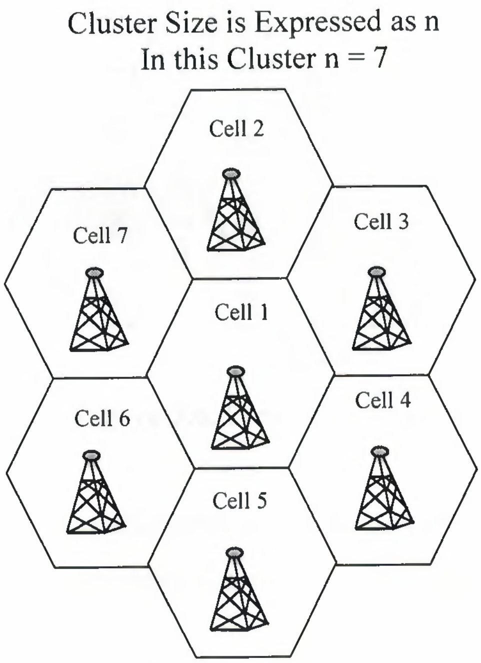

Figure 3.5. A Seven-Cell Cluster

3.4.3 Frequency Reuse

Because only a small number of radio channel frequencies were available for mobile systems, engineers had to find a way to reuse radio channels to carry more than one conversation at a time. The solution the industry adopted was called frequency planning or frequency reuse. Frequency reuse was implemented by restructuring the mobile telephone system architecture into the cellular concept.

The concept of frequency reuse is based on assigning to each cell a group of radio channels used within a small geographic area. Cells are assigned a group of channels that is completely different from neighboring cells. The coverage area of cells is called the footprint. This footprint is limited by a boundary so that the same group of channels can be used in different cells that are far enough away from each other so that their frequencies do not interfere.

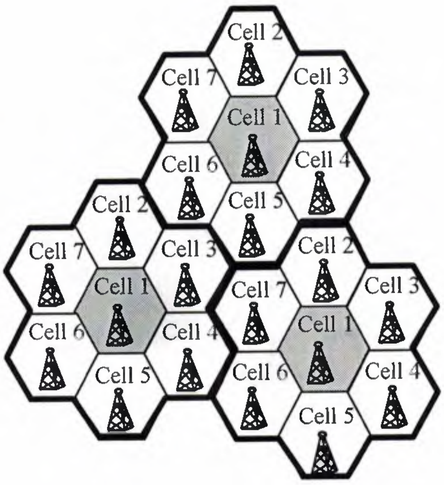

Figure 3.6. Frequency Reuse

Cells with the same number have the same set of frequencies. Here, because the number of available frequencies is 7, the frequency reuse factor is 1/7. That is, each cell is using 1/7 of available cellular channels.

3.4.4 Cell Splitting

Unfortunately, economic considerations made the concept of creating full systems with many small areas impractical. To overcome this difficulty, system operators developed the idea of cell splitting. As a service area becomes full of users, this approach is used to split a single area into smaller ones. In this way, urban centers can be split into as many areas as necessary to provide acceptable service levels in heavy-traffic regions, while larger, less expensive cells can be used to cover remote rural regions.

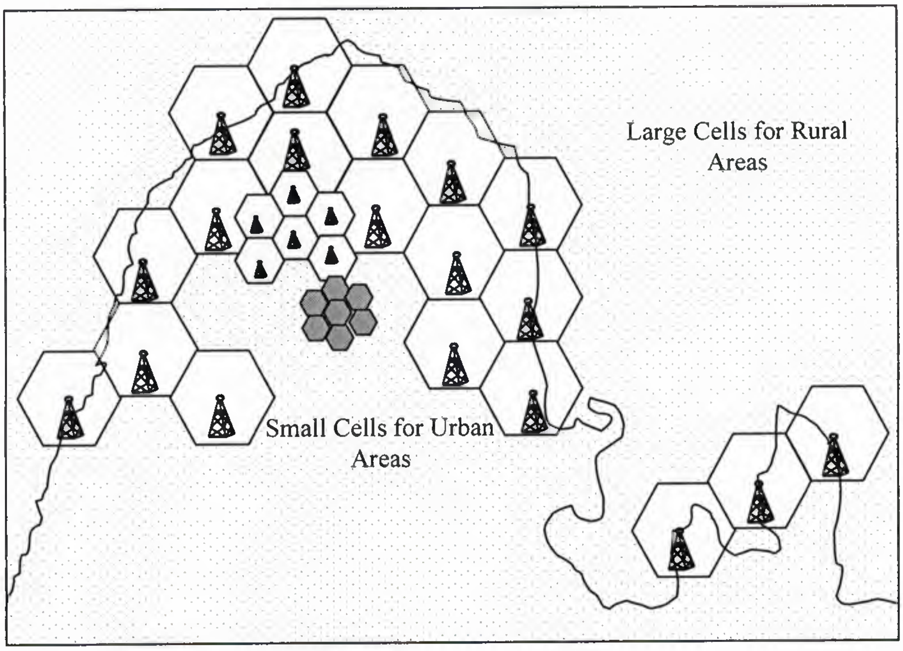

Figure 3.6. Cell Splitting

3.4.5 Handoff

The final obstacle in the development of the cellular network involved the problem created when a mobile subscriber traveled from one cell to another during a call. As adjacent areas do not use the same radio channels, a call must either be dropped or transferred from one radio channel to another when a user crosses the line between adjacent cells. Because dropping the call is unacceptable, the process of handoff was created. Handoff occurs when the mobile telephone network automatically transfers a call from radio channel to radio channel as mobile crosses adjacent cells.

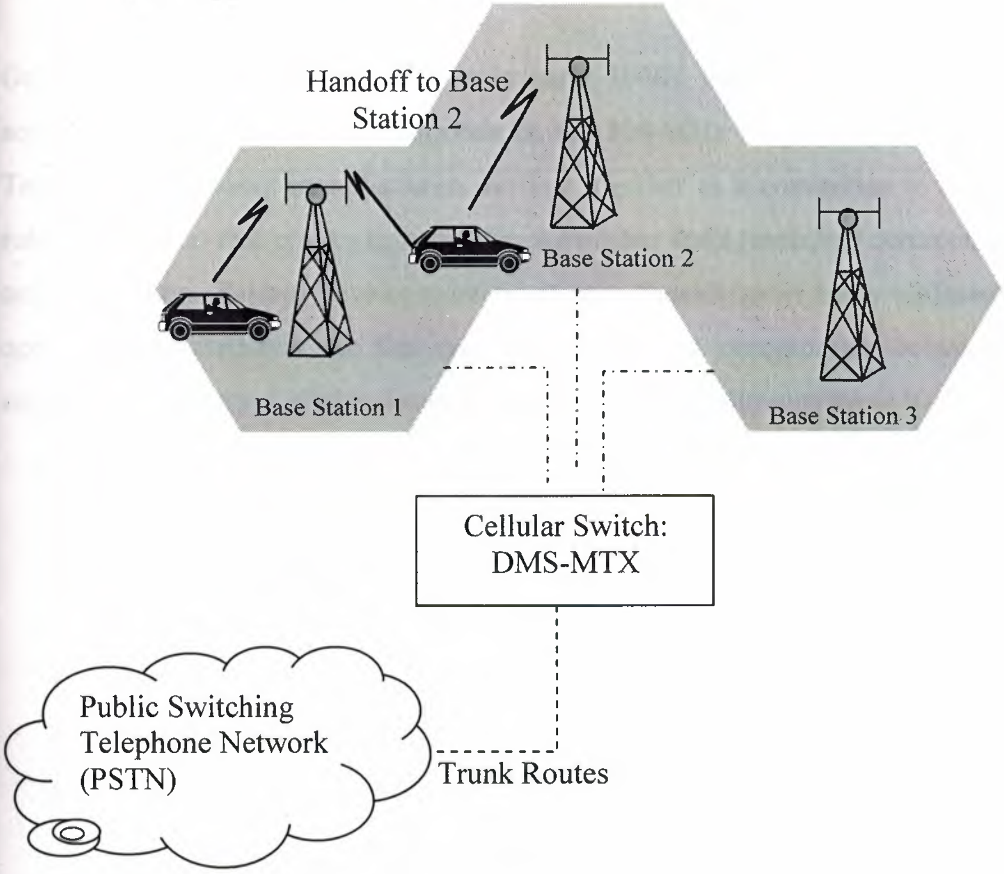

Figure 3.7. Handoff between Adjacent Cells

During a call, two parties are on one voice channel. When the mobile unit moves out of the coverage area of a given cell site, the reception becomes weak. At this point, the cell site in use requests a handoff. The system switches the call to a stronger-frequency channel in a new site without interrupting the call or alerting the user. The call continues as long as the user is talking, and the user does not notice the handoff at all.

3.5 North American Analog Cellular Systems

Originally devised in the late 1970s to early 1980s, analog systems have been revised somewhat since that time and operate in the 800-MHz range. A group of government, Telco, and equipment manufacturers worked together as a committee to develop a set of rules (protocols) that govern how cellular subscriber units (mobiles) communicate with the cellular system. System development takes into consideration many different, and often opposing, requirements for the system, and often a compromise between conflicting requirements results. Cellular development involves the following basic topics:

- frequency and channel assignments
- type of radio modulation
- maximum power levels
- modulation parameters
- messaging protocols
- call-processing sequences

3.5.1 The Advanced Mobile Phone Service (AMPS)

AMPS was released in 1983 using the 800-MHz to 900-MHz frequency band and the 30-kHz bandwidth for each channel as a fully automated mobile telephone service. It was the first standardized cellular service in the world and is currently the most widely used standard for cellular communications. Designed for use in cities, AMPS later expanded to rural areas. It maximized the cellular concept of frequency reuse by reducing radio power output. The AMPS telephones (or handsets) have the familiar telephone-style user interface and are compatible with any AMPS base station. This makes mobility between service providers (roaming) simpler for subscribers. Limitations associated with AMPS include the following:

- low calling capacity
- limited spectrum
- no room for spectrum growth
- poor data communications
- minimal privacy
- inadequate fraud protection

AMPS is used throughout the world and is particularly popular in the United States, South America, China, and Australia. AMPS uses Frequency Modulation (FM) for radio transmission. In the United States, transmissions from mobile to cell site use separate frequencies from the base station to the mobile subscriber.

3.5.2 Narrowband Analog Mobile Phone Service (NAMPS)

Since analog cellular was developed, systems have been implemented extensively throughout the world as first-generation cellular technology. In the second generation of analog cellular systems, NAMPS was designed to solve the problem of low calling capacity. NAMPS is now operational in 35 U.S. and overseas markets, and NAMPS was introduced as an interim solution to capacity problems. NAMPS is a U.S. cellular radio system that combines existing voice processing with digital signaling, tripling the capacity of today's AMPS systems. The NAMPS concept uses frequency division to get 3 channels in the AMPS 30-kHz single channel bandwidth. NAMPS provides 3 users in an AMPS channel by dividing the 30-kHz AMPS bandwidth into 3-10 kHz channels. This increases the possibility of interference because channel bandwidth is reduced.

3.6 Cellular System Components

The cellular system offers mobile and portable telephone stations the same service provided fixed stations over conventional wired loops. It has the capacity to serve tens of thousands of subscribers in a major metropolitan area. The cellular communications system consists of the following four major components that work together to provide mobile service to subscribers.

- public switched telephone network (PSTN)
- mobile telephone switching office (MTSO)
- cell site with antenna system
- mobile subscriber unit (MSU)

3.6.1 PSTN

The PSTN is made up of local networks, the exchange area networks, and the long-haul network that interconnect telephones and other communication devices on a worldwide basis.

3.6.2 Mobile Telephone Switching Office (MTSO)

The MTSO is the central office for mobile switching. It houses the mobile switching center (MSC), field monitoring, and relay stations for switching calls from cell sites to wire line central offices (PSTN). In analog cellular networks, the MSC controls the system operation. The MSC controls calls, tracks billing information, and locates cellular subscribers.

3.6.3 The Cell Site

The term *cell site* is used to refer to the physical location of radio equipment that provides coverage within a cell. A list of hardware located at a cell site includes power sources, interface equipment, radio frequency transmitters and receivers, and antenna systems.

3.6.4 Mobile Subscriber Units (MSUs)

The mobile subscriber unit consists of a control unit and a transceiver that transmits and receives radio transmissions to and from a cell site. The following three types of MSUs are available:

- the mobile telephone (typical transmit power is 4.0 watts)
- the portable (typical transmit power is 0.6 watts)
- the transportable (typical transmit power is 1.6 watts)

The mobile telephone is installed in the trunk of a car, and the handset is installed in a convenient location to the driver. Portable and transportable telephones are hand-held and can be used anywhere. The use of portable and transportable telephones is limited to the charge life of the internal battery.

3.7 Digital Systems

As demand for mobile telephone service has increased, service providers found that basic engineering assumptions borrowed from wire line (landline) networks did not hold true in mobile systems. While the average landline phone call lasts at least 10 minutes, mobile calls usually run 90 seconds. Engineers who expected to assign 50 or more mobile phones to the same radio channel found that by doing so they increased the probability that a user would not get dial tone—this is known as call-blocking probability. As a consequence, the early systems quickly became saturated, and the quality of service decreased rapidly. The critical problem was capacity. The general characteristics of time division multiple access (TDMA), Global System for Mobile Communications (GSM), personal communications service (PCS) 1900, and code division multiple access (CDMA) promise to significantly increase the efficiency of cellular telephone systems to allow a greater number of simultaneous conversations. Figure 3.8 shows the components of a typical digital cellular system.

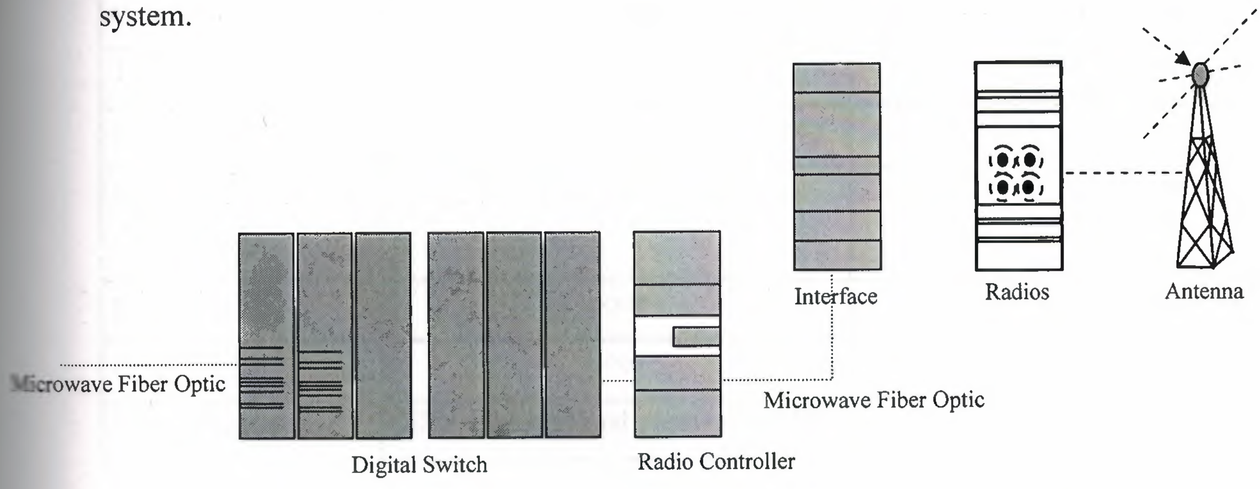

Figure 3.8. Digital Cellular System

The advantages of digital cellular technologies over analog cellular networks include increased capacity and security. Technology options such as TDMA and CDMA offer more channels in the same analog cellular bandwidth and encrypted voice and data. Because of the enormous amount of money that service providers have invested in AMPS hardware and software, providers look for a migration from AMPS to digital analog mobile phone service (DAMPS) by overlaying their existing networks with TDMA architectures.

Table 3.1. AMPS/DAMPS Comparison

	Analog	**Digital**
standard	EIA–553 (AMPS)	IS–54 (TDMA + AMPS)
spectrum	824 MHz to 891 MHz	824 MHz to 891 MHz
channel bandwidth	30 kHz	30 kHz
channels	21 CC/395 VC	21 CC / 395 VC
Conversations per channel	1	3 or 6
subscriber capacity	40 to 50 conversations per cell	125 to 300 conversations per cell
TX/RCV type	continuous	time shared bursts
carrier type	constant phase variable frequency	constant frequency variable phase
mobile/base relationship	mobile slaved to base	authority shared cooperatively
privacy	poor	better—easily scrambled
noise immunity	poor	high
fraud detection	ESN plus optional password (PIN)	ESN plus optional password (PIN)

3.7.1 Time Division Multiple Access (TDMA)

North American digital cellular (NADC) is called DAMPS and TDMA. Because AMPS preceded digital cellular systems, DAMPS uses the same setup protocols as analog AMPS. TDMA has the following characteristics:

1. IS–54 standard specifies traffic on digital voice channels
2. initial implementation triples the calling capacity of AMPS systems
3. capacity improvements of 6 to 15 times that of AMPS are possible
4. many blocks of spectrum in 800 MHz and 1900 MHz are used
5. all transmissions are digital
6. TDMA/FDMA application 7. 3 callers per radio carrier (6 callers on half rate later), providing 3 times the AMPS capacity

TDMA is one of several technologies used in wireless communications. TDMA provides each call with time slots so that several calls can occupy one bandwidth. Each caller is assigned a specific time slot. In some cellular systems, digital packets of information are sent during each time slot and reassembled by the receiving equipment into the original voice components. TDMA uses the same frequency band and channel allocations as AMPS. Like NAMPS, TDMA provides three to six time channels in the same bandwidth as a single AMPS channel. Unlike NAMPS, digital systems have the means to compress the spectrum used to transmit voice information by compressing idle time and redundancy of normal speech. TDMA is the digital standard and has 30-kHz bandwidth. Using digital voice encoders, TDMA is able to use up to six channels in the same bandwidth where AMPS uses one channel.

3.7.2 Extended Time Division Multiple Access (E–TDMA)

The E–TDMA standard claims a capacity of fifteen times that of analog cellular systems. This capacity is achieved by compressing quiet time during conversations. E–TDMA divides the finite number of cellular frequencies into more time slots than TDMA. This allows the system to support more simultaneous cellular calls.

3.7.3 Fixed Wireless Access (FWA)

FWA is a radio-based local exchange service in which telephone service is provided by common carriers. It is primarily a rural application—that is, it reduces the cost of conventional wire line. FWA extends telephone service to rural areas by replacing a wire line local loop with radio communications. Other labels for wireless access include fixed loop, fixed radio access, wireless telephony, radio loop, fixed wireless, radio access, and Ionic. FWA systems employ TDMA or CDMA access technologies.

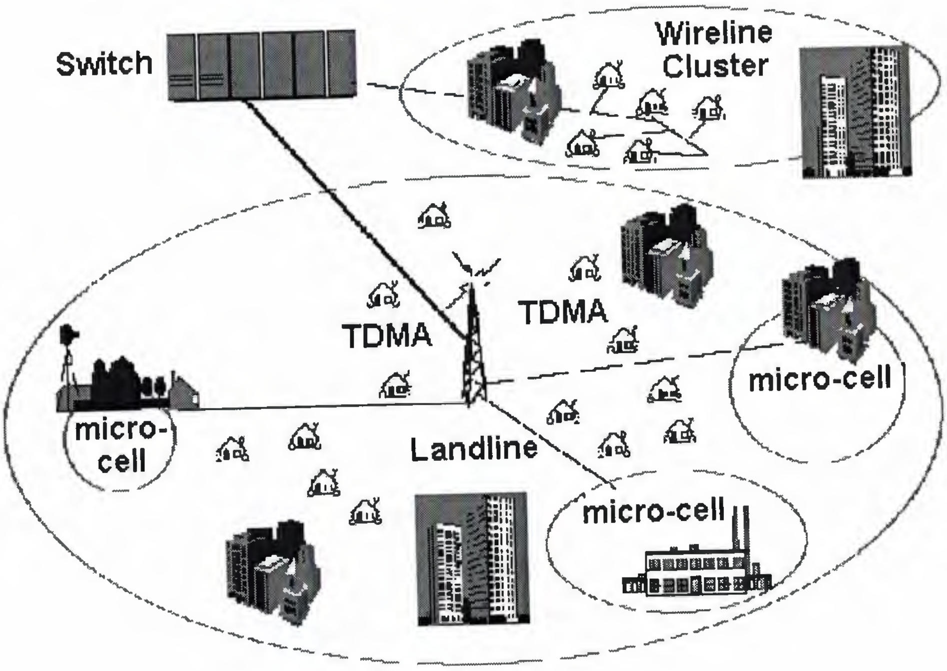

Figure 3.9 Fixed Wireless Access

3.7.4 Personal Communications Service (PCS)

The future of telecommunications includes PCS. PCS at 1900 MHz (PCS 1900) is the North American implementation of digital cellular system (DCS) 1800 (GSM). Trial networks were operational in the United States by 1993, and in 1994 the Federal Communications Commission (FCC) began spectrum auctions. As of 1995, the FCC auctioned commercial licenses. In the PCS frequency spectrum, the operator's authorized frequency block contains a definite number of channels. The frequency plan assigns specific channels to specific cells, following a reuse pattern that restarts with each *n*th cell. The uplink and downlink bands are paired mirror images. As with AMPS, a channel number implies one uplink and one downlink frequency (e.g., Channel 512 = 1850.2-MHz uplink paired with 1930.2-MHz downlink).

3.7.5 Code Division Multiple Access (CDMA)

CDMA is a digital air interface standard, claiming 8 to 15 times the capacity of analog. It employs a commercial adaptation of military, spread-spectrum, single-sideband technology. Based on spread spectrum theory, it is essentially the same as wire line service-the primary difference is that access to the Local Exchange Carrier (LEC) is provided via wireless phone. Because users are isolated by code, they can share the same carrier frequency, eliminating the frequency reuse problem encountered in AMPS and DAMPS. Every CDMA cell site can use the same 1.25-MHz band, so with respect to clusters, $n = 1$. This greatly simplifies frequency planning in a fully CDMA environment.

CDMA is an interference-limited system. Unlike AMPS/TDMA, CDMA has a soft capacity limit; however, each user is a noise source on the shared channel and the noise contributed by users accumulates. This creates a practical limit to how many users a system will sustain. Mobiles that transmit excessive power increase interference to other mobiles. For CDMA, precise power control of mobiles is critical in maximizing the system's capacity and increasing battery life of the mobiles. The goal is to keep each mobile at the absolute minimum power level that is necessary to ensure acceptable service quality. Ideally, the power received at the base station from each mobile should be the same (minimum signal to interference).

4. BLUETOOTH

4.1 Overview

Bluetooth is a standardized technology that is used to create temporary (ad-hoc) short-range wireless communication systems. These Bluetooth wireless personal area networks (WPAN) are used to connect personal accessories such as headsets, keyboards, and portable devices to communications equipment and networks. Bluetooth was named after Harald Blatand, King of Denmark. King Blatand was head of Denmark from 940 to 985 A.D and he is known for uniting the Danes and Norweigans. It seems appropriate to name the wireless technology that unifies communication between diverse sets of devices after King Blatant

Figure 4.1 shows the different types of devices that can be linked by wireless personal area network communication. This example shows that the computer can be located near the devices such as a keyboard, mouse, display, speakers, microphone, and a presentation projector. As these devices are brought within a few feet of each other, they automatically discover the Availability and capabilities of other devices. If these devices have been setup to allow communication with other devices, the user will be able to use these devices as if they were directly connected with each other. As the devices are removed from the area or turned off, the option to use these devices will be disabled from the user.

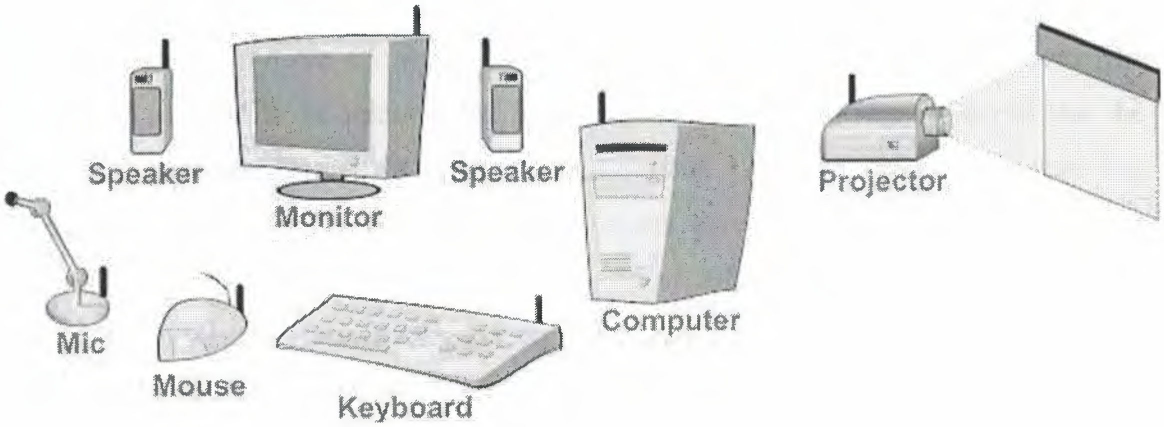

Figure 4.1 Wireless Personal Area Network Devices

The Bluetooth system operates in 2.4 GHz unlicensed (uncontrolled) frequency bands with very low radio transmission power of 1 mill watt to 100 mill watts. For unlicensed use, radio transmission is authorized for all users provided the radio equipment conforms to unlicensed requirements. Anyone can use the unlicensed frequency band but there is no guarantee they will perform at peak performance due to possible interference. Devices within the frequency bands are required to operate in such a way that they can co-exist in the same area with minimal interference with each other. While users are not required to obtain a license to use devices that operate in unlicensed frequency bands, the manufacturers of devices are required to conform to government regulations. These regulations also vary from country to country.

4.2 Development Timeline

Bluetooth evolved from simple replacement of wires to a dynamically changing wireless personal area network (WPAN). Bluetooth was first conceived in 1993 at Ericsson as a way to allow portable cellular telephones to get smaller while user devices such as PDAs and Laptops could interface with communication devices without the need for wires. In 1988, several companies setup agreements to form the special interest group (SIG) to develop and promote Bluetooth technology. In July 1999, version 1.0 of the Bluetooth specification was published. In December 1999, Additional promoter companies were added and a revised specification 1.0B was released. In 2001, Bluetooth specification 1.1 was released and in November 2003, Bluetooth specification 1.2 was released. Figure 4.2 shows how Bluetooth evolved from the time it was first conceived in 1993 to its status at the beginning of 2004. This diagram shows that the Bluetooth SIG was formed in 1998 and that version 1.0 Bluetooth specification was released in mid-1999. The first Bluetooth products were qualified in June 2000. There were 83 qualified Bluetooth products by the end of 2000. In February 2001, Bluetooth specification version 1.1 was released and there were 481 qualified products by the end of 2001. An additional 412 Bluetooth products were qualified in 2002 brining the total number of qualified Bluetooth products to 893 at the end of 2002. In November 2003, Bluetooth specification 1.2 was released and the total number of Bluetooth qualified products that were available at the end of 2003 was 1336.

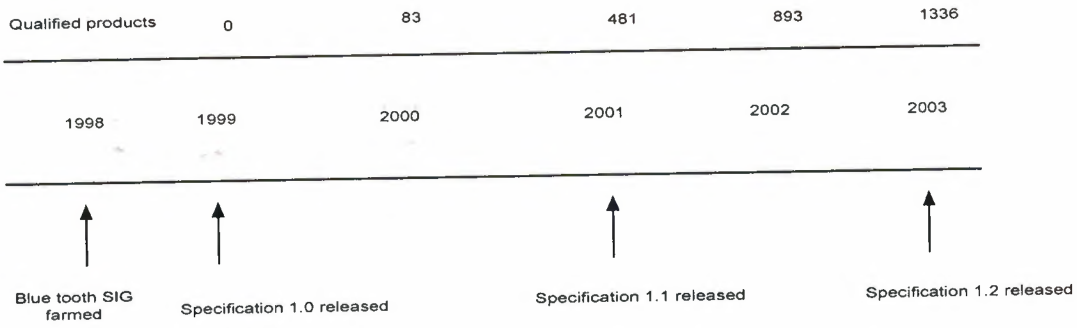
Qualified products
0
83
481
893
1336
1998
1999
2000
2001
2002
2003
Blue tooth SIG farmed
Specification 1.0 released
Specification 1.1 released
Specification 1.2 released

Figure 4.2 Bluetooth Timeline

4.3 Special Interest Group (SIG)

A special interest group works to help develop and promote information about a specific technology, product, or service. The Bluetooth SIG oversees the certification of Bluetooth assemblies and devices and promotes the use of Bluetooth technology. To allow the SIG to achieve its objectives, the SIG has setup or recognizes specific groups or facilities to be part of the development process. These include the Bluetooth qualification body (BQB), Bluetooth qualified test facility (BQTF), Bluetooth qualification review board (BQRB), and the Bluetooth qualification administrator (BQA).

The BQB is authorized by the Bluetooth qualification review board (BQRB) to be responsible for the checking of declarations and documents against requirements, reviewing product test reports, and listing conforming products in the official database of Bluetooth qualified products. BQTFs are test labs that are recognized and certified by the BQRB as being capable of testing and qualifying Bluetooth devices. The BQRB is responsible for managing, reviewing, and improving the Bluetooth qualification program through which vendor products are tested for conformance. The Bluetooth qualification administrator (BQA) is responsible for overseeing the administration of the qualification program.

The BQA ensures that qualified products can be listed on the Bluetooth website. In addition to the test programs, testing events ("UnplugFests") are made available to members to allow compatibility testing. UnplugFests (UPF) are testing events in which manufacturers or developers agree to test their products with other products in a secret closed environment. There are approximately 3 UnplugFests per year.

The participation in UnplugFests allows manufacturers and developers to find problems with their products or areas of correction or clarification that are needed in the Bluetooth specifications. Figure 4.3 shows the general product qualification process used by the Bluetooth special interest group (SIG) to ensure reliable operation and compatibility between Bluetooth devices. This example shows that the first step for product qualification is for the company to gather the program reference documents (PRDs), complete the product documentation, and submit the documents to the Bluetooth Qualification Board (BQB). The company then develops the prototype of the product and submits the product to a Bluetooth Qualification Test Facility (BQTF) for testing. The BQTF test report is then sent to the Bluetooth Qualification Board (BQB) for review. If the product documentation and test results are accepted by the BQB, the product will be added to the Bluetooth qualified product list.

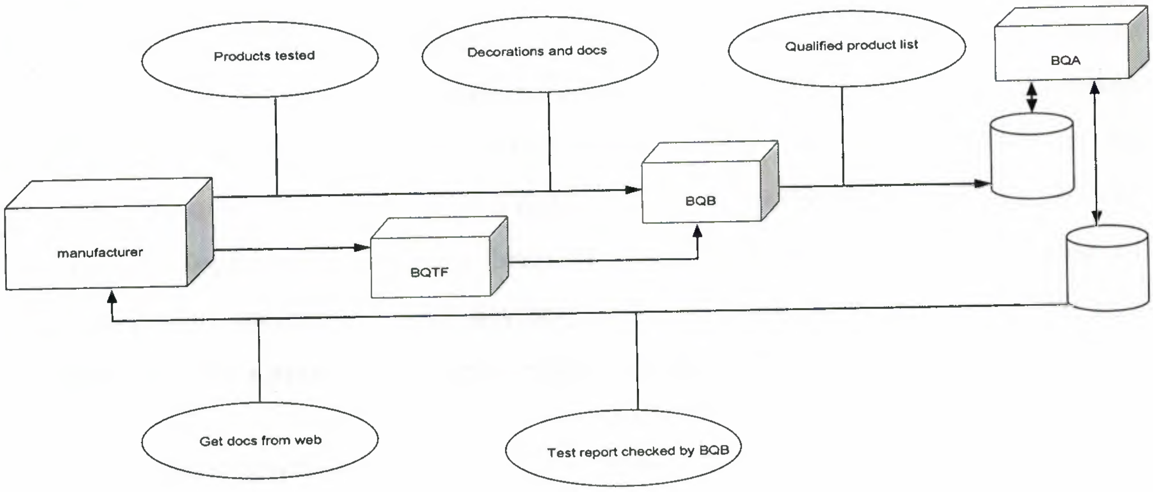

Figure 4.3 Bluetooth product qualification process

4.4 Bluetooth Basics

Bluetooth is a wireless personal area network (WPAN) communication system Standard that allows for wireless data connections to be dynamically added and removed between nearby devices. Each Bluetooth wireless network can contain up to 8 active devices and is called a Piconet. Piconets can be linked to each other (overlap) to form larger area Scatternets. The system control for Bluetooth requires one device to operate as the coordinating Device (a master) and all the other devices are slaves. This is very similar to the structure of a universal serial bus (USB) system that is commonly used in personal computers and devices such as digital cameras. However, unlike USB connections, most Bluetooth devices can operate as either a master (coordinator) or slave (follower) and Bluetooth devices can reverse their roles if necessary. The characteristics of Bluetooth include an unlicensed frequency band that ranges from 2.4 GHz to 2.483 GHz. This frequency band was chosen because it is available for use in most countries throughout the world. While the standard frequency band for Bluetooth is in the 2.400 GHz to 2.483 GHz (83 MHz) frequency band, the original Bluetooth specification had an optional smaller frequency band 23 MHz version for use in some countries. The use of a smaller frequency band does not change the data transmission rate, however, these devices will be more sensitive to interference (such as other Bluetooth device transmission) and this interference may cause a lower overall data transmission rate.

Every Bluetooth device has a unique 48-bit address BD_ADDR (pronounced "B-D-Adder"). In addition to identifying each Bluetooth device, this address is used to determine the frequency hopping pattern that is used by the Bluetooth device. Bluetooth devices may have different power classification levels. The 3 power versions for Bluetooth include; 1 mW (class 3), 2.5 mW (class 2) and 100 mWatts (class 1). Devices that have an extremely low power level of 1 mill watt have a very short range of approximately 1 meter. Bluetooth devices that have a power level of up to 100 mill watts can provide a transmission range of approximately 100 meters.

The high power version (class 1) is required to use adjustable (dynamic) power control that automatically is reduced when enough signal strength is available between Bluetooth devices. Because the objectives of Bluetooth are low power and low complexity, the simple modulation type of Gaussian frequency shift keying (GFSK) is used. This modulation technology represents a logical 1 or 0 with a shift of 115 kHz above or below the carrier signal. The data transmission rate of the RF channel is 1 Mbps.The smallest packet size in the Bluetooth system is the Bluetooth packet data unit (PDU). Bluetooth PDUs are transmitted between master and slave devices within a Bluetooth Piconet. Each PDU contains the address code of the Piconet, device identifier, and a payload of data. When the PDUs are used to carry logical channels, part of the data payload includes a header, which includes logical channel identifiers. The length of the PDU can vary to fit within 1, 3 or 5 time slot period (625 use per time slot). Control message PDUs (e.g. link control) always fit within 1 time slot system), device identifier (specific device within the piconet), logical channel identifiers (to identify ports), and a payload of data. If a specific protocol is used (such as a wireless RS-232 communication port - RFCOMM), an additional protocol service multiplexer (PSM) field is included at the beginning of the payload data. This diagram also shows that the PDU size can have a 1, 3 or 5 slot length.

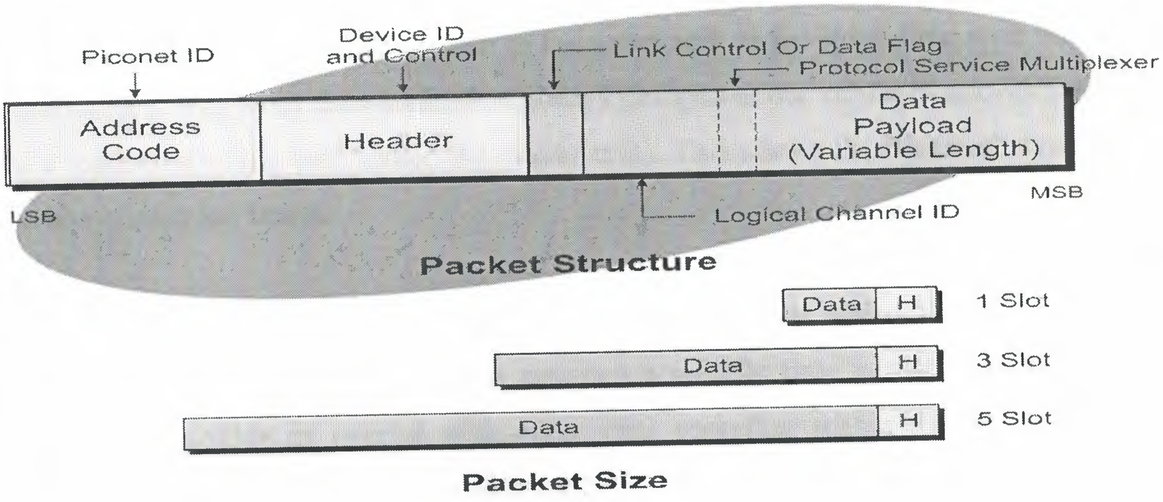

Figure 4.4 Bluetooth Packet Structure

The Bluetooth system uses time division duplex (TDD) operation. TDD operation permits devices to transmit in either direction, but not at the same time the basic radio transmission process used in the Bluetooth system. This diagram shows that the frequency range of the Bluetooth system ranges from 2.4 GHz to 2.483 GHz and that the basic radio transmission packet time slot is 625 usec. It also shows that one device in a Bluetooth piconet is the master (controller) and other devices are slaves to the master. Each radio packet contains a local area piconet ID, device ID, and logical channel identifier.

This diagram also shows that the hopping sequence is normally determined by the master's Bluetooth device address. However, when a device is not under control of the master, it does not know what hopping sequence to use, it listens for inquiries on a standard hopping sequence and then listens for pages using its own Bluetooth device address.

4.5 Temporary Small Networks (Piconets)

Bluetooth forms temporary small networks of Bluetooth communication devices of up to 8 active devices called Piconets. The Bluetooth system allows for wireless data connections within the Piconet to be dynamically added and removed between nearby devices. Because the Bluetooth system hops over 79 channels, the probability of interfering (overlapping) with another Bluetooth system is less than 1.5%. This allows several Bluetooth Piconets to operate in the same area at the same time with minimal interference. Bluetooth communication always designates one of the Bluetooth devices as a main controlling unit (called the master unit). This allows the Bluetooth system to be non-contention based.

This means that after a Bluetooth device has been added to the temporary network (the Piconet), each device is assigned a specific time period to transmit and they do not collide or overlap with other units operating within the same Piconet. Multiple Piconets can be linked to each other to form Scatternets. Scatternets allow the master in one Piconet to operate as a slave in another Piconet. While this allows Bluetooth devices in one Piconet to communicate with devices in another Piconet (cross-Piconet communication), the use of Scatternets require synchronization (and sharing of data transmission Bandwidth) making them inefficient.

4.6 Data Transmission Rates

The basic (gross) radio channel data transmission rate for a single Bluetooth radio channel is 1 Mbps with over 723.2 kbps available to a single user. The data rate available to each user is less than the radio channel data transmission rate because some of the data transmission is used for control and channel management purposes. The users in each Piconet split the remaining data transmission rate. Bluetooth Piconet that provides for headset operation, which uses 64 kbps channels in both directions, uses a total data transmission rate of 128 kbps. This is approximately 25% of the total available data transmission bandwidth. The Bluetooth system allows for different rates in different directions (asynchronous) or for equal data rate (symmetrical rate) transmission.

Figure 4.5 shows how the radio channel data transmission rate for Bluetooth devices is divided between transmission directions and between multiple devices. In example 1, a PDA is transferring a large file to a laptop computer using asymmetrical transmission. During the transfer, it uses the 5-slot packet size to reach the maximum data transmission rate of 723.2 kbps from the PDA to the laptop. This only allows a data transmission rate of 57.6 kbps from the laptop to the PDA. Example 2 shows a symmetrical data transmission rate of 433.9 kbps between two video conferencing stations. How the data transmission rate from a laptop is shared between a wireless headset and a PDA. This example shows that the headset uses a symmetrical data transmission rate of 64 kbps from device 1 (the master coordinator) to device 2 and a 57.6 kbps asynchronous data transmission rate between the Laptop and the PDA.

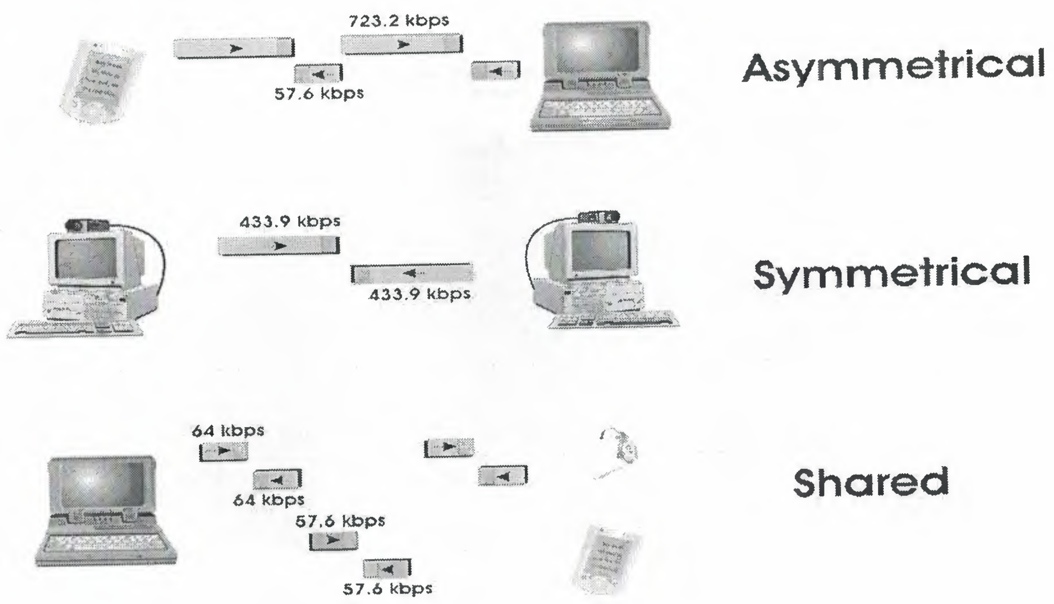

Figure 4.5 Bluetooth Data Transmission Rates

4.7 Frequency Hopping Spread Spectrum (FHSS)

Frequency hopping spread spectrum (FHSS) is a radio transmission process where a message or voice communication is sent on a radio channel that regularly changes frequency (hops) according to a predetermined code. The receiver of the message or voice information must also receive on the same frequencies using the same frequency hopping sequence. Frequency hopping was first used for military electronic countermeasures. Because radio communication occurs only for brief periods on a radio channel and the frequency hop channel numbers are only known to authorized receivers of the information, transmitted signals that use frequency hopping are difficult to Detect and monitor. Figure 4.6 shows a simplified diagram of how the Bluetooth system uses frequency hopping to transfer information (data) from a transmitter to a receiver using 79 communication channels. This diagram shows a transmitter that has a preprogrammed frequency tuning sequence and this frequency sequence occurs by hopping from channel frequency to channel frequency. To receive information from the transmitter, the receiver uses the exact same hopping sequence. When the transmitter and receiver frequency hopping sequences occur exactly at the same time, information can transfer from the transmitter to the receiver. This diagram shows that after the transmitter hops to a new frequency, it transmits a burst of information (packet of data). Because the receiver hops to the same frequency, it can receive the packet of data each time.

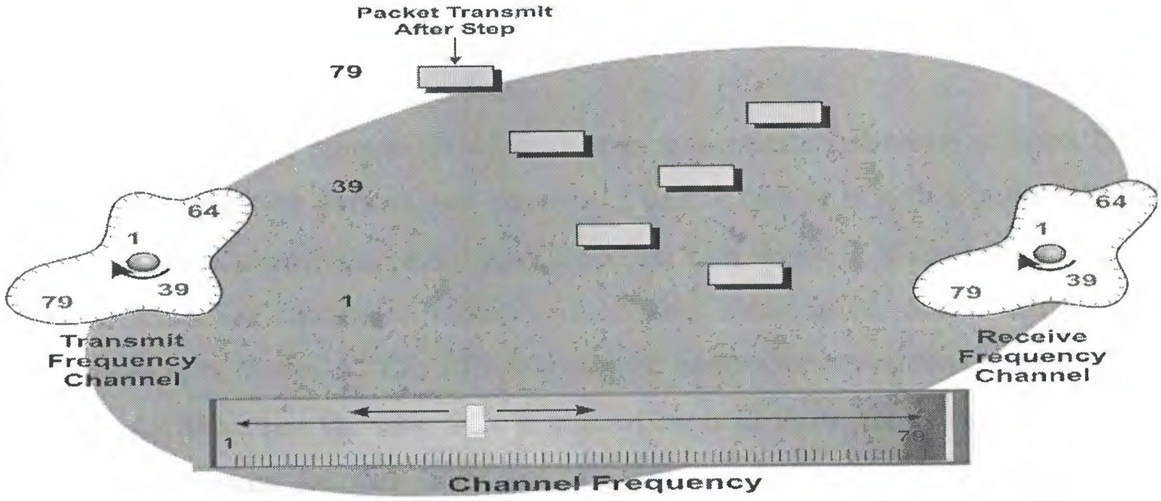

Figure 4.6 Bluetooth Frequency Hopping Operation

4.8 Service Discovery

Service discovery is the process of finding other devices that can communicate with your device and determining what capabilities they have that you may want to use. Service discovery protocol (SDP) is the communication messaging protocol used by a communication system (such as the Bluetooth system) to allow devices to discover the availability and capabilities of other nearby devices. The SDP process is similar to a registry in Windows as it dynamically creates a list of available resources. The discovery process begins with an inquiry message that a device sends that can be received by nearby devices. These devices are constantly looking for an inquiry message to respond to. When a device receives an inquiry message, it responds with an address that can be used to establish a connection with the device. If a device wants to discover the services of another device, it must use the device address and establish a temporary connection. The name and capabilities of the device can be discovered using service discovery protocol (SDP). The discovery process is optional. Devices can be programmed not to respond to inquiry messages.

4.9 Pairing with Other Bluetooth Devices

Bluetooth pairing is an initialization procedure whereby two devices communicating for the first time create an initial secret link key that will be used for subsequent authentication. For first-time connection, pairing requires the user to enter a Bluetooth security code or PIN. Since neither unit knows any secret keys of the other unit, a new secret key must be created. As a result, both devices request a PIN to be entered. Both users must enter the same PIN. This PIN is then used with other information to create a secret key in each unit

The Bluetooth pairing process allows devices to authenticate and create a secure link between two Bluetooth devices. For example, if user A desires to push a business card to user B, user A attempts to establish a connection with device B. If device B has been setup to reject the connection unless a device has been paired, device B requests authentication of device A. This PIN is combined with other information to produce a secret key.) Assuming the same PIN is entered by both users, the secret key that is created is the same for both users and this key can be used to help authenticate (validate) the identify of the other user.

4.10 Bonding with Other Bluetooth Devices

Bonding is the process of creating a very secure link key that is shared between Bluetooth devices. One way to create a very secure link key is to use the initial secret link key that was created by the pairing process. The bonding process involves encoding (modifying) a unique unit key (created when the Bluetooth device is turned on) using the initial secret link key and sending the unit key to the other unit. This allows the Bluetooth devices to use each other's secret information to create a more secure link key. This secure link key can be used in future authentication validations.

4.11 Connecting with Other Bluetooth Devices

Connecting Bluetooth devices is the process of creating a communication session between devices. Communication sessions are the end-to-end transmission links between devices during operation of a software program or logical connection between two communications devices. In communications s ystems, the session involves the establishment of a physical channel, logical channel(s).The configuration of transmission parameters operation of higher-level applications and termination of the session as the application is complete. During a session, many processes or message transmissions may occur. Creating a connection between Bluetooth devices involves getting the attention of a device through paging and allowing the device to change it's frequency hopping sequence so that it can become part of the Piconet. If the Bluetooth device knows the address of the Bluetooth device it wants to connect to, the connection process usually takes less than 1-2 seconds. The ability of a device to allow "connections" is optional. Bluetooth devices can be programmed to not allow other devices to connect to them ("non-connectable.")

The connection process begins with the master unit changing its hopping sequence to the hopping sequence of the recipient device. The Bluetooth unit first sends many identification (ID) packets alerting the receiving device that someone wants to connect to them. When the receiving device (the slave) hears its ID address, it can immediately respond as no other units will be competing for its own access code. When the master hears that the recipient has responded, it sends a frequency hopping synchronization (FHS) packet that contains its Bluetooth Address. Both the master and slave then change their hopping sequence to the master's hopping sequence (the Piconet address). The master will then send a Poll message to the slave using the new hopping sequence and if the slave responds (usually with a null-no information response), the master knows it is connected to the recipient (slave) device. Figure 4.7 shows how a Bluetooth device can connect to other devices. This example shows that the Bluetooth master unit first sends many ID packets using the hopping sequence of the recipient device. When the receiving device hears its ID address, it immediately responds. This allows the master to send a FHS packet that contains the masters Bluetooth Address. Now both the master and slave change their hopping sequence to the Piconet hopping sequence (determined by the master's Bluetooth Address).

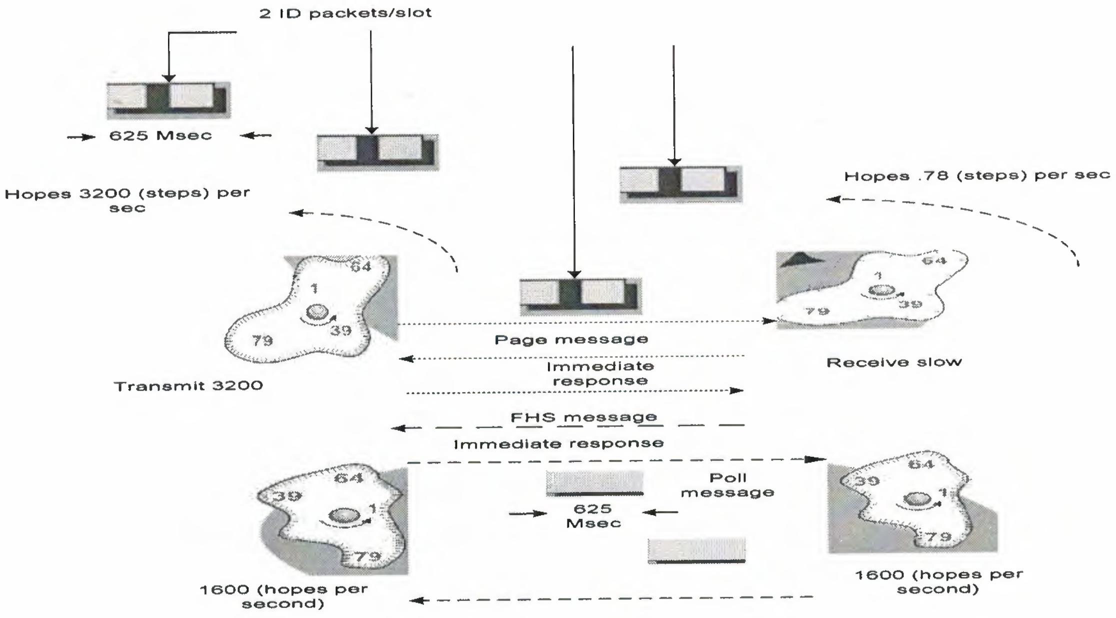

Figure 4.7 Bluetooth Connection Operations

4.12 Summary

Bluetooth covers the basics of short-range Bluetooth wireless networks used by accessories such as headsets, keyboards, cameras, and printers. This chapter includes operational descriptions of Bluetooth and home RF and how these technologies are evolving.

5. GENERAL HCI BRIDGING CONSIDERATION FOR BLUETOOTH

5.1 Overview

This chapter gives an overview of Bluetooth wireless technology and how Xilinx high-volume programmable devices can be used to integrate Bluetooth network interfaces at the system level. The Xilinx device families targeted at these high-volume applications include XC9500 CPLDs and Spartan FPGAs.

The flow of this document will be to start with an overview of Bluetooth. We will next examine the major functional blocks of a Bluetooth interface and give an overview of the Application Specific Standard Products (ASSPs) that are available to implement them. We will then illustrate the Host Controller Interface (HCI) that is standard for interfacing the Bluetooth subsystem to the host.

While this document focuses on the use of these devices in Bluetooth HCI interface applications, the examples discussed illustrate many of the issues found in other designs, specifically, how to cost effectively interface complex ASSPs with incompatible interfaces. The ASIC vendors have abandoned the traditional solution for this class of problems, the small ASIC, as they moved towards the system on chip market. Fortunately for system designers, new classes of low cost PLDs, such as the Spartan family, have filled this void with devices that replace low density ASICs and retain the time to market advantages of FPGAs.

5.2 Bluetooth Background

Bluetooth is a short-range radio link that is intended to replace cabling used to connect fixed or portable electronic devices. Bluetooth devices operate in the unlicensed 2.4 GHz Industrial, Scientific, and Medical (ISM) frequency band. The Bluetooth baseband protocol supports both circuit switched and packet-switched communications and uses Frequency Hopping Spread Spectrum (FHSS) technology for transmission. In North

America and most of Euro, Bluetooth operates in the frequency range from 2.402 to 2.480 GHz, with this band divided into 79, 1 MHz sub-channels.

Up to eight Bluetooth devices with overlapping coverage share channel bandwidth and form what is called a piconet. On each piconet one Bluetooth unit acts as the master while the other unit(s) acts as slave(s). In addition to the seven slaves that may be active on each piconet, more slaves can remain locked to the master in a parked state. The piconet master controls channel access for both active and parked slaves.

Multiple piconets with overlapping coverage areas form a scatternet. While each piconet can only have a single master, slaves can participate in different piconets on a time-division basis.

In addition, a master in one piconet can be a slave in another piconet. The Bluetooth protocol includes support for both packet and circuit switching. Each piconet can support an asynchronous data channel, up to three simultaneous synchronous voice channels, or a mix of links. Each voice channel is 64 kb/s synchronous, full duplex, and is referred to as a Synchronous Connection-Oriented (SCO) link. The asynchronous channels are called Asynchronous Connection-Less (ACL) links and can support an asymmetric link of maximally 723.2 kb/s in either direction while permitting 57.6 kb/s in the return direction, or a 433.9 kb/s symmetric link.

The Bluetooth specification breaks the functions required to implement a Bluetooth interface into three major functional blocks as shown in Figure 5.1. These functional blocks map directly to the partitioning of the Bluetooth specification.

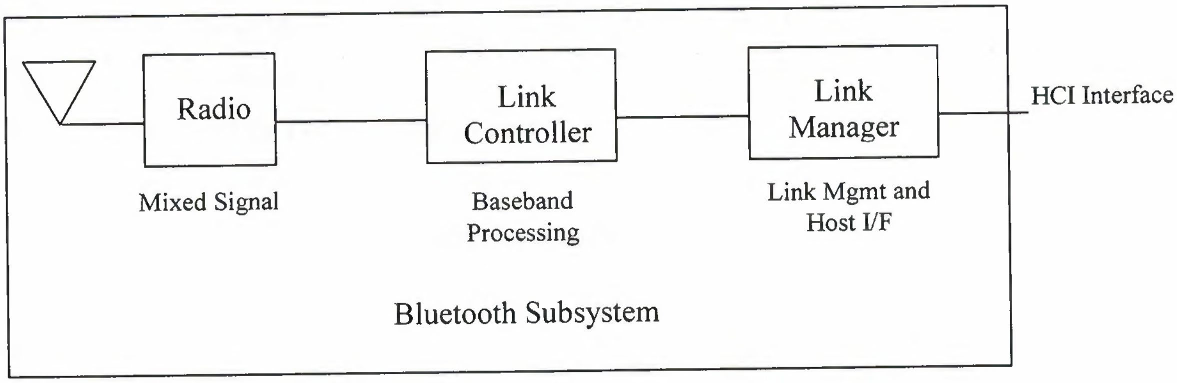

Figure 5.1 Bluetooth Functional Blocks

5.2.1 Bluetooth Radio

The Radio implements the broadband air interface for Bluetooth devices. The radio is typically implemented as a multi-chip module and includes an antenna switch, baluns, amplifiers, digital PLL based clock recovery, modulation, and demodulation circuitry.

5.2.2 Bluetooth Link Controller

The Link Controller Bluetooth baseband functions are described in Part B of the specification, and consist of low-level link layer functions. Baseband functions include:

- CVSD speech coding
- Header Error Check (HEC) generation and checking
- Forward Error Correction (FEC) generation and checking
- Cyclic Redundancy Check (CRC) generation and checking
- Data whitening (scrambling)
- Payload encryption and decryption
- Sequencing of frequency hopping

5.2.3 Bluetooth Link Manager

The Link Management block implements the Link Manager Protocol (LMP), which handles low level control plane functions such as:

- Link setup between devices
- Generating, exchanging, and checking encryption keys
- Negotiation of baseband packet sizes
- Power modes and duty cycles of the radio
- Connection states of the unit in a piconet

The complexity of these functions mandates a software implementation, typically running on an embedded RISC processor. This software approach leads to the use of the processor for other functions as well, including the firmware required for interfacing to the host system.

5.3 The Bluetooth Host Controller Interface

The Bluetooth specification defines the Host Controller Interface (HCI) as follows:

"The HCI provides a command interface to the baseband controller and link manager, and access to hardware status and control registers. This interface provides a uniform method of accessing the Bluetooth baseband capabilities."

The HCI consists of two parts — the software that implements the command interface and the physical hardware that is used to connect the Bluetooth subsystem to the host. The purpose of the HCI software is to make the hardware that comprise the interface transparent to higher level software in the system.

5.3.1 HCI Software

The Bluetooth software architecture consists of two types of components — Data Plane and Control Plane. Data Plane components are responsible for the transfer of data across the link.

Control Plane components are responsible for link control and management. For the purposes of this chapter we will be focused on the Data and Control Plane components that make up the HCI. Figure 5.2 illustrates the HCI software architecture and how it relates to the Bluetooth host interface hardware.

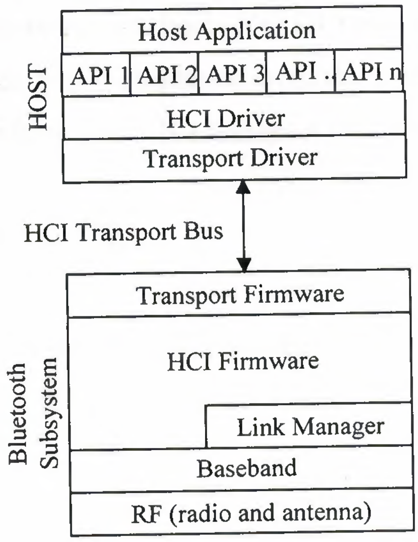

Figure 5.2 Bluetooth HCI Software Architecture

5.3.2 HCI Commands and Events

The host controls the Bluetooth network interface through a variety of commands that are provided by the HCI driver. In addition to these commands, the Bluetooth specification defines a set of events that are generated by the HCI firmware in the Bluetooth Network Interface to indicate state changes in the interface.

HCI commands and events are combined with data from ACL and SCO links over the interface hardware that is used for HCI transport. The scheme used to multiplex this data over the interface is specific to each interface. Figure 5.3 illustrates how this works.

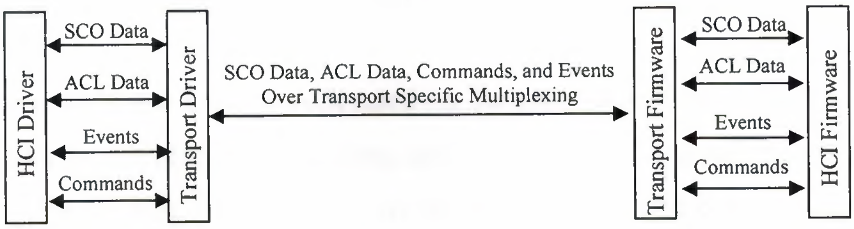

Figure 5.3 HCI Multiplexing

5.3.3 HCI Hardware/Transports

The HCI transports define how to transfer three classes of data between the Bluetooth network interface and the Bluetooth host. The HCI transports define how each of these data types is encapsulated and multiplexed over the interface hardware. The Bluetooth specification currently defines three HCI transport layers.

- UART Transport Layer
- RS232 Transport Layer
- USB Transport Layer

A white chapter from the Bluetooth SIG describes a fourth, the PC Card Transport Layer. An overview of each of these transport layers follows.

5.3.4 UART Transport Layer

Both the RS232 and the UART transport layers use serial communication via a UART to transfer data between the Bluetooth Network Interface and the Bluetooth Host. The difference between the two lies in the environmental assumptions that drive the design of the protocol.

The UART transport layer is designed for an environment where the Bluetooth Network Interface and the host are located on the same printed circuit board and, as a result, the link is relatively error free. The encapsulation of the HCI commands consists of an HCI Packet Indicator, which indicates whether the frame contains a command, an event, or a data packet followed by a length indicator. The length indicator is used to check whether frame synchronization has been lost and a simple recovery mechanism is used in the event that it is.

Since both ends of the link are collocated on the same board, the UART transport layer does not specify the electrical signaling and in most cases this will be done using TTL levels. Since both ends of the link are on the same board, there is no mechanism defined for baud rate negotiation.

5.3.5 RS232 Transport Layer

The RS232 transport layer was designed to support communication between a Bluetooth Network Interface and a host located in separate enclosures. As a result, RS232 electrical signaling is specified and a far more elaborate link protocol is defined. Both of these are needed in order to deal with a link that spans larger distances and must be able to handle a much higher line error rate. In addition to the four HCI Packet Indicators used by the UART transport layer, the RS232 layer adds two additional types. One is used to negotiate the baud rate, parity type, number of stop bits, and the protocol mode. The second is used to communicate line errors to the transmitter. The RS232 transport layer defines two encapsulation schemes. One uses HDLC like framing and a 16-bit CRC. The second scheme o mits the CRC and uses the RTS/CTS signals for delineation. Both schemes include a sequence number in each frame so that receivers can easily detect lost frames.

5.3.6 USB Transport Layer

Unlike the UART and RS232 transports, the endpoint mechanisms defined in USB provide a simple means of multiplexing traffic over the link without additional overhead to identify the type of traffic. As a result, this specification primarily describes how the Bluetooth data types are mapped to USB endpoints. This includes how to map SCO data streams to the USB isochronous data services.

5.3.7 PC Card Transport Layer

This layer is not defined in the Bluetooth specification, but is described in a white chapter. The reason for this is that the Bluetooth specification decided not to restrict the implementation details other than the requirement that the card comply with the requirements of the PC Card and Cardbus standards. In order to support interoperability, the manufacturer must provide an interface hardware specific transport driver that can be used with the HCI driver on the host system.

5.3.8 HCI Support in ASSPs

Unless you have the time and resources to implement a Bluetooth network interface from scratch, you will need to base your HCI implementation strategy on what is provided by existing ASSPs. Table 5.1 provides a survey of the available ASSPs that implement Bluetooth Link Controller and/or Link Manager functionality.

Table 5.1 Bluetooth ASSPs

Vendor	Part Number	Radio	Link Controller	Link Manager	Processor	HCI Transport Support
Qualcomm	MSM3300		√	√	ARM7TDMI	USB, UART
Ericsson	ROK 101 007		√	√	ARMT-Thumb	USB, UART
National Semiconductor	LMX5001		√		None	None
Silicon Wave	SiW1601		√		None	None
Texas Instruments	BSN6030		√	√	ARM7TDMI	UART
Infineon	BlueMoon 1	√	√	√	TBD	UART
Lucent	W7400		√	√	TBD	USB, UART
Alcatel	TBD	√	√	√	ARM7	UART
Cambridge Silicon Radio	BlueCore	√	√	√	Proprietary	USB, UART
Philsar Semiconductor	PH2410		√	√	ARM7TDMI	USB, UART

For the purposes of our further discussion, these ASSPs fall into two broad categories:1) those that implement only the Link Controller functions and 2) those that include both the Link Controller and the Link Manager. The key difference between these two classes of devices is that the first includes an embedded RISC processor used to implement Link Manager and HCI functions and the second does not.

The decision regarding which class device to use requires analyzing the tradeoffs for each alternative. In the next section we will review these tradeoffs and show how Spartan devices can provide solutions to HCI interface issues.

5.4 The Challenges of Creating Real World Solutions

The cost effective integration of Bluetooth technology into a system level design can present many challenges, including:

- Evolving standards. While the Bluetooth core protocols are stable, the mapping of higher level protocols and services such as IP traffic is still evolving.
- Buggy ASSPs. Most of the ASSPs that are being used are relatively new and will have deficiencies that must be dealt with by the system designer.
- Emerging product use models. In most cases Bluetooth technology is currently too expensive for it's original target market, cable replacement. This and the lure of other potential "killer apps" has driven many to push Bluetooth into application areas that are still being defined.

Clearly what is needed is a flexible technology that allows system designers to quickly develop Bluetooth solutions, but at the same time can respond to this fluid environment. Let's explore how Spartan devices can meet this need when used to interface the two classes of Bluetooth ASSPs that we have discussed.

5.4.1 Link Controller + Link Manager ASSPs

If you choose to use one of the ASSPs that implements both Link Controller and Link Manager functions you have the advantage of starting with a fairly complete solution. The manufacturers of these devices provide not only a RISC processor and one or more of the defined HCI transport interfaces, but also provide HCI and Link Manager firmware. With these devices, system level integration consists of connecting the HCI transport interface to the host. If the host system has an available USB or UART port, system level integration can be accomplished with no additional hardware. In many cases, these ports are integrated into the system's core logic chipset. Figure 5.4 illustrates a system that utilizes this type of zero-glue interface.

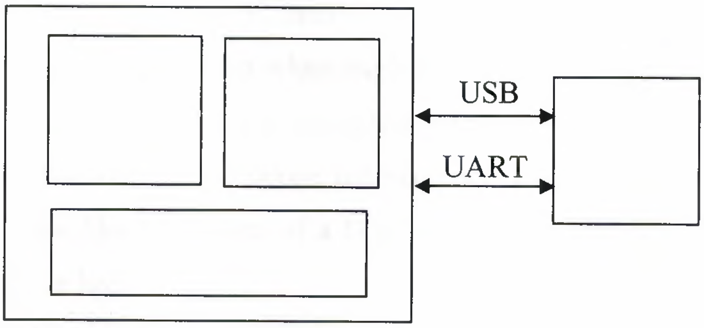

Figure 5.4 Zero-Glue Bluetooth Interface

In many cases this approach may not be workable for any of the following reasons:

- The host system may not have sufficient USB or UART ports for the application. This can occur when the host has a limited number of ports, or when that application requires a large number of ports.
- The host system may not have serial ports that can support the data rate required for full Bluetooth performance. In order to support maximum system level performance, Bluetooth ASSPs include UARTs that are capable of supporting transfer rates of up to 1.5 Mbps.
- Standard port interface ASSPs are not tailored for Bluetooth protocol handling, and as a result, they may consume significant processing resources when operating at Bluetooth

rates. As we will see, interrupt processing overhead can become a drain on host processing resources.

In any of these situations a Spartan device can be used to implement the required interface hardware. In deeply embedded applications this would typically be an interface between a USB or UART core and the local bus of the embedded processor used in the system. While Spartan devices can be used to implement either a USB or UART transport interface, the simpler transport protocol of the UART interface results in a more cost effective interface with better system level performance.

A UART interface is more cost effective since it requires fewer hardware resources to implement. Since the UART only operates in a single mode, eight bits of data, no parity, and one stop bit, the implementation can be very simple and operate at a very high speed. In addition, unlike the USB or RS232 interfaces, the UART transport layer does not need external level shifters or transceivers when implemented in an FPGA.

Better system level performance is a key advantages of Spartan FPGAs. They can be used to quickly create interface solutions tailored specifically for the target application. Figure 5.5 shows the block diagram of a UART enhanced with DMA and HCI frame transfer state machine logic.

This design is described in detail in XAPPxxx.

The use of Spartan FPGAs improve system level performance by reducing the overhead of servicing interrupts for transmission and reception. Unlike traditional UARTs where an interrupt is generated every time the small on-chip FIFOs are filled or emptied, this design generates an interrupt only when a complete HCI frame has been transmitted or received. This is accomplished by having application specific logic that decodes the frame size information in the HCI header and configures the DMA logic appropriately. This logic also checks to ensure that proper frame level synchronization is being maintained. The net result being that the burden of interrupt handling is considerably reduced for the host processor. As a result, more processor performance is available for other value added functions. In addition, with the wide range of standard interface and memory controller IP available, the Spartan device can be used to implement other core logic functions as well.

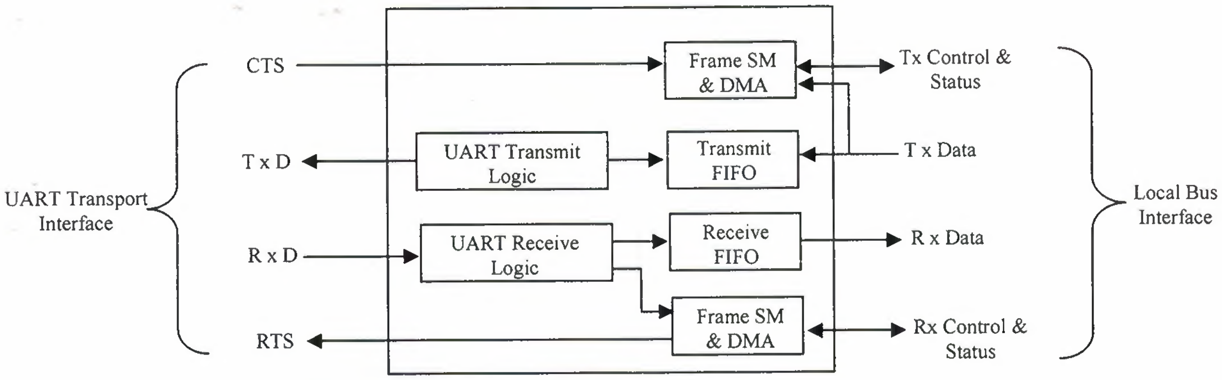

Figure 5.5 Bluetooth Application Tailored UART

5.4.2 Link Controller Only ASSPs

Link controller logic only ASSPs can result in additional design work in order to create a complete solution. They also present the opportunity for tighter integration and lower system cost for deeply embedded systems. This is due to the fact that if the Bluetooth Network Interface and the host are located on the same board, there is no need for an HCI transport layer. In this case, the interfaces used to control the Baseband Processing and Radio functions, as well as the interfaces used to transfer data frames, are simply interfaced directly to the host processor. Since these interfaces are usually specific to the ASSP involved, a Spartan device provides a low cost means of interfacing them to either the host processors bus or an I/O bus such as PCI (Figure 5.6). Note that here is another situation where the Spartan device can be used to integrate other core logic functions as well.

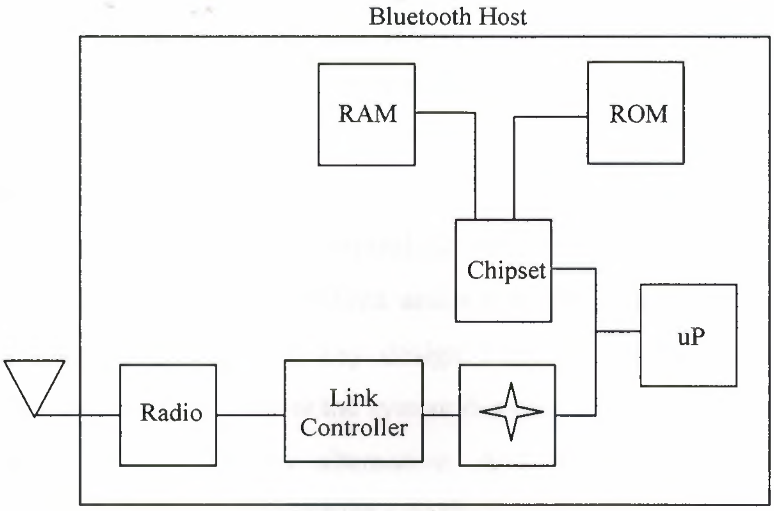

Figure 5. 6 Link Controller ASSP Interface

In this arrangement, the host CPU takes on all protocol processing functions. These results in further economies since the RISC processor that was dedicated to Link Management functions and its non-volatile memory requirements are eliminated from the system. In this architecture, the Bluetooth software stack collapses to the arrangement shown in Figure 5.7.

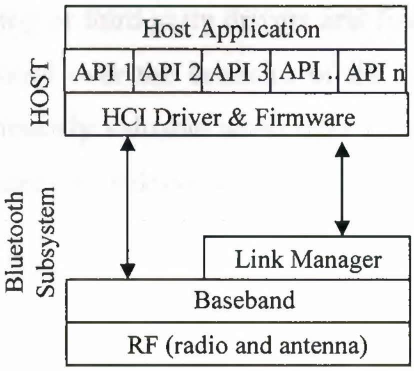

Figure 5.7 Collapsed Software Stack

\s we have seen, programmable logic provides an excellent platform for integrating 3luetooth technology into embedded systems. Let's highlight the key benefits that they ›ring:

.4.3 Time-to-Market

Xilinx programmable logic provides several advantages that reduces time-to-market. First, ι broad range of IP support from Xilinx and a ever growing number of third party IP ›artners provide quick access to key design building blocks. Second, as benchtop ›rogrammable solutions, they allow the system designer to achieve a functional hardware ›latform more rapidly than any alternative. And third, programmable devices are ;tandard parts that are easy to reproduce quickly in limited-to-high volumes to capture a ›osition in strategic accounts before the competition.

.4.4 Rapid Software Development

Software development is one of the biggest issues in integrating Bluetooth technology. \nd, obviously, since programmable logic can achieve functional hardware sooner it :reates an advantage in this area. However, this advantage can be even greater when you :onsider the flexibility that programmability brings to the equation. For instance, it is ›ften desirable to use existing or third party drivers and firmware. With a programmable solution, you have full control over the behavior of the interface ensuring a workable ıpproach. This can be particularly valuable when third party code is involved because it nay not be well documented or understood and modifications can raise support and naintenance concerns.

5.4.5 Time-in-Market

Product development by its nature is not an exact science. Bugs and incompatibilities are simply a reality that engineering must deal with. Here, especially Xilinx programmable devices can provide a valuable advantage, as our solutions are inherently **reprogrammable**. Thus, patches for known problems can be put into production as soon as they are validated on the existing hardware revision and can also be deployed to installed systems. This allows you to keep your existing design shipping and greatly reduces the risk of obsolete part inventories and expensive field replacement programs.

5.4.6 Rapid Design Derivation

A system design is a corporate asset and in today's world of hyper competition and compressed development cycles, these assets must be flexible. Standards evolve, customers request new features, and experience reveals new business opportunities that can be exploited. Thoughtful designs that incorporate programmable logic are inherently more scalable and are superior platforms for rapid and efficient product derivation. Thus, well exploited programmable logic can make your future product roadmap a strategic competitive advantage.

5.4.7 System Level Cost Reduction

In the past, the use of programmable logic was considered an expensive solution. However, times have changed because Moore's Law has worked to the advantage of programmable solutions. Today, $10 will buy 100,000 system gates in volume, off the shelf, and ready to go. And, as these devices usually replace other functions in your design as well, they can often enable real system level cost reductions. Programmable logic has replaced the small cost reduction ASICs of yesterday and brings many other advantages to your system too!

CONCLUSION

As this brief introduction to mobile networking has shown, Mobile IP has great potential. Security needs are getting active attention and will benefit from the deployment efforts underway. Within the IETF, Mobile IP is likely to move from a proposed standard to a draft standard in the near future.

The IETF standardization process requires the working group to rigorously demonstrate interoperability among various independent implementations before the protocol can advance. FTP Software has hosted two interoperability testing sessions, and many vendors have taken advantage of the opportunity. Test results have given added confidence that the Mobile IP specification is sound, implementable, and of diverse interest throughout the Internet community. Only a few minor revisions have been needed to ensure the specification can be interpreted in only one way by the network protocol engineers and programmers who must implement it.

It is possible that the deployment pace of Mobile IP will track that of IPv6, or that the requirements for supporting mobility in IPv6 nodes will give additional impetus to the deployment of both IPv6 and mobile networking. The increased user convenience and the reduced need for application awareness of mobility can be a major driving force for adoption. Since both IPv6 and Mobile IP have little direct effect on the operating systems of mobile computers outside of the network layer of the protocol stack, application designers should find this to be an acceptable programming environment. Of course, everything depends heavily on the willingness of platform and router vendors to implement Mobile IP and/or IPv6, but indications are strong that most major vendors already have implementations either finished or underway.

REFERENCES

] Vineet Sachdev, System Engineer

] Trueposition Inc. 1111 West DeKalb Pike Wayne, PA 19087

] Asha Mehrotra, "GSM System Engineering"

] MOBILE COMMUNICATIONS SERIES, Artech House Publishers.

] Brian McIntosh "Telecommunications"
tp://telecomindustry.about.com/business/telecomindustry/library/weekly/aa1115999.ht

] Dick Tracy "The Applications We Promote…." http://www.comm-nav.com/commnav.htr

] GRAYSON WIRELESS, a division of Allen Telecom. "Geometrix Wireless Location
ensor" http://java.grayson.com/geodatasheet.htrr

] Louis A. Stilp "Examining the Coming Revolution in Location Services"
tp://www.trueposition.com

] Paul J, Bouchard "AccuCom Wireless Service Inc." http://www.Global-Images.com

0] Tutorial "How GPS works?" http://www.trimble.com