Gateway security between Bluetooth and GSM/GPRS

Master Thesis
in
Information and Communication
Technology

By
Kjetil Jørgensen Wiig

Waterford Institute of Technology
May, 2002

Abstract

The amount of wireless technologies is getting bigger and wider. It is more and more common in everyday use for a huge amount of the people. A cellular phone is something that everyone got in addition to normal access the Internet. So, why not both at the same time. When data is transmitted over the Internet and are communicating with wireless technologies, there are several vulnerable points. There have to be secure data communication, and the security is only something that should lay behind which people do not have to think about. Unfortunately, this is not the way it is, there are different technologies that cover the security better than others.

This master thesis covers some of the different technologies that is in everyday use. Where the concentration is on the wireless technologies and how to combine these securely. The technologies cowered are; Bluetooth, GSM/GPRS, IPSec and WTLS covered by WAP.

One essential purpose of WAP is to provide a way to bring commercial applications into mobile terminals. Security mechanisms of GSM or GPRS do not necessarily provide enough security for all WAP applications. Additional security mechanisms need to be implemented to the mobile terminals and the network. Mobile terminals unfortunately usually have limited computing resources, and implementing efficient security mechanism can be difficult.

Since both WAP enabled GSM/GPRS telephones and Bluetooth devices support the use of WAP and IP communications, it is nothing that says that WTLS and IPSec could not been used to secure the communication.

Both, WTLS and IPSec support some kind of end-to-end security. This thesis covers some of the solutions that may be used together with the wireless technologies GSM/GPRS and Bluetooth. In the end there are an overview over different combinations and a suggestion to secure the data communication.

Preface

This thesis is a part of the Master of Engineering degree in Information and Communication Technology at Agder University College in Grimstad. The assignment is the final work on the education that leads to the Norwegian degree Sivilingeniør, which is equivalent to the Master of Science degree. It is performed by one or two students who work with the given task, which is given by the college or by the industry. The workload of the assignment is supposed to be similar to the workload of one semester.

This assignment is given by Agder University College in Grimstad in co-operation with Waterford Institute of Technology, Ireland. It gives a theoretical evaluation of different security solutions for use between a GPRS mobile and a Bluetooth network thru the public communication.

I will especially thank my supervisor, Dr Paul O'Leary at Waterford Institute of Technology for counselling and support throughout the report.
Thanks go also Magne Arild Haglund and Lars Line at Agder University College that made this thesis aboard to real, and to my mates Kjell Rune Øyrås and Håkon Gunleifsen for advices and discussions. At the end a very special thank to my wife Camilla Grefstad Wiig that has been supporting and with me the whole time aboard.

Waterford
27. May, 2002

_____

Kjetil Jørgensen Wiig

Contents

List of figures

1 Introduction

1.1 Background

We are moving towards a society with great demands for the future. People would like to have the same facilities when they are travelling as they have at the office and at home. The last years of revolutionary services for the Internet and especially for associated services has given the possibility to access information sources worldwide.

The number of terminals with wireless access to network resources will increase, as people are more on the move, and with the demand for network access.

It is important to take care of the security and protect the subscribers and the communication they may have, even if this is useless information or financial transactions that others want to harm or take advantage of.

Most of the communication technologies have underlying mechanisms to take care of the security. The subscriber's does not even have to think about it. The general security is always a minimum, especially in mobile devices that have limited processing power and smaller bandwidth.

This master thesis will guide you through different wireless technologies and evaluate the security they cover. It will also give you some combinations of security solutions and recommendations that may be used when communicating with a GSM/GPRS telephone or a Bluetooth device on the Internet.

1.2 The Thesis Definition

Gateway security between Bluetooth and GSM/GPRS

Bluetooth is an emerging technology for short range wireless communication which enables devices within a radius of 10 meters to communicate with each other. This technology may also be used in customer premises for control and monitoring of devices. For example, intelligent homes can control the heating, light, video recorder or even the garage door. Or small businesses can use stock control or credit verification from a remote till.

If you connect these devices together with a Bluetooth module with GSM or GPRS capability, it is possible to implement a number of control and/or monitoring tasks. However, clearly this may have security implications as others could also gain control of these operations.

This problem has two aspects, the authentication of the user and the subsequent integrity of communications. The aim of the project would be to examine both of these security aspects when GSM/GPRS is used as a gateway to control devices on a Bluetooth network.

In the Bluetooth Generic Access Profile (GAP), 3 Security modes are defined:

- Security Mode 1: non-secure
- Security Mode 2: service level enforced security
- Security Mode 3: link level enforced security

However, the interface and access security from a WAN technology such as GSM/GPRS have not yet been clearly defined. The project should examine the existing Bluetooth security, the security required for the interface and the level of security attainable using GSM/GPRS to access the Bluetooth network. The project should also discuss the merits of different possible solutions, evaluate and discuss security solutions and if possible suggest improvements for some applications.

1.3 Method

This assignment was more an evaluation than a research project. The work has involved reading and absorption a lot of information. Most of this material is found on the Internet, such as; Technical Specifications, White Papers and other papers in addition to books.

In the beginning of the project, there were set up a studying schedule that contained the main technologies and the amount of time on each. As the project progressed the schedule was changed slowly but not drastically, and always together with my supervisor Dr Paul O'Leary.

The outline of this thesis is as follows:
First, a look at different security definitions that has to be aware of, followed by a description of problems in distributed systems.
Bluetooth in general and more specific on the different levels of security for both devices and services. The key management, authentication and encryption processes, and ends up with a small evaluation.
GPRS in general and the security mechanisms it provides. The authentication algorithm and some vulnerabilities in 2G Security.
IPSec with the AH and ESP in both the Tunnel and Transport Mode. Key management and some weaknesses.
WAP with the concentration on WTLS followed by known security holes.
End to end security based on WAP transactions, with different solutions.
Then the conclusion, which starts with a smaller conclusion on the different technologies, and ends with combining them.

There has been a guidance meeting with the supervisor every week during this thesis project.

2 Definitions and Distributed Systems Security

In this section, the project take a brief look into the general security definitions and definitions of threats to computer systems, problems with distributed systems and a look at security in Ad Hoc networks.

2.1 Definitions

Confidentiality, Integrity and Authentication, normally referred to as CIA, are three different services that computer and network security should cover. In addition to CIA there are; access control, non-repudiation and Denial-of-Service which should be considered when designing a communication-system: [11]

- Confidentiality
  - Normally associated with user data confidentiality.
  - One definition is that the used transfer method ensures a private end-to-end transfer. This means the property of information has not been disclosed to unauthorised parties. This involves the leakage of information from the system to a part that should not have seen it.
  - A plaintext is simply encrypted and decrypted to implement confidentiality. If the plaintext is encrypted using a strong encryption, it is almost impossible for eavesdropper to decrypt and read the original content.
  - It has traditionally been seen as the most formidable problem in communications systems.
  - The requirements of the strong encryption are met when the security is created by using a shared secret, not secret algorithm. The key space, from which the used shared secret is chosen, has to be large. Moreover, the used cryptographic method must produce an output which appears random to all statistical tests. And the used method should be resistant to all known attacks.
  - However, encrypted data is useless if the recipient is not able to decrypt data. The sender and the recipient have to share a method to encrypt and decrypt the data. They both have to know the used cryptographic method and the shared secret. The shared secret is a piece of information known by both parties but nobody else.
  - There is also other kind of privacy. It is not always the case that the information has a recipient. Sometimes there is data that is not supposed to be decrypted by anybody e.g. Unix-type passwords. This kind of encryption method is called a one way encryption, this means that there is no formula for a reverse process back to the original information. Hash-values are the most common method for the one way encryption.

- Integrity
  - Normally associated with error correction and retransmission techniques to ensure that data is not corrupted. The property of information that has not been changed by unauthorised parties.
  - Applies to techniques to ensure unauthorised change of the information, that data isn't wilfully modified by an adversary. Cryptographical checksum of the original information is a technique to ensure that data has not been modified. Just a plain checksum is not enough. There is need for some sender-related information mixed into calculations, e.g. information is signed with the user's digital signature.
  - In most cases, integrity is more critical than guaranteeing privacy. It is more important that the information is received unaltered but seen by someone else, than somebody has been able to modify it without making out the whole information.
  - May include sequence- and flow control for packet switching.

- Authentication
  - Normally used in reference to the verification of the user identity. In other words; it is a technique to ensure that the stated identity of the user is correct.
  - Also applies to data origin verification.
  - The contacting party has to present some verification to prove its identity. Challenge-Response is a common authentication mechanism that actively challenges the user to prove that he is the right person, and then the user has to give the right response.
  - Authentication may be one-way or mutual.
  - After the authentication, the service provider can be sure that the service is available to the user who has correct rights to use the service. On the other hand, the user can be confident about the service provider.

- Access control
  - Access control is about giving authorised users access to a service while denying unauthorised users the same service.
  - A popular technique is to use an Access Control List (ACL) for specifying who should have access (and possibly also specifying further details of user rights).
  - Access control depends on proper authentication.

- Denial-of-Service
  - While access control in many ways can be seen as a security service to deny unauthorised users access to services, Denial-of-Service can be seen as a security service to ensure that authorised users are not denied access to services.
  - Normally this threat involves that an access to a system resource is being blocked by a malicious attacker, and is a threat to the availability of the system

- Denial-of-Service also includes physical access, which means that downtime of all sorts is a problem for a Denial-of-Service requirement.

- Non-repudiation
  - Non-repudiation is all about the sender and/or the receiver being accountable for what they have sent/received; That is, the user cannot deny having sent a message and the receiver cannot deny having received the message.
  - The non-repudiation service is often associated with Digital Signatures and Trusted Third Party (TTP).
  - Non-repudiation is hard to provide and is not normally provided in full.
  - Financial transaction systems must provide a trustworthy non-repudiation service (and some WAP services (pay-by-phone) may fall into this category).
  - Non-repudiation assumes cryptographical data integrity.

2.2 Problems with distributed systems

In distributed systems, objects are scattered in different places. This makes the security issues more difficult than regular centralised computer systems.

One problem is that there are usually several parties involved. There may not be a clear consensus on the security policy and if different participants enforce different kinds of security policies, collaboration is impossible.

Another problem is a process called delegation. When a user logs in remotely to a network and wants to execute a program on a remote machine, some problems arise. The user needs certain rights to use the resources on the remote machine. The user typically delegates his access rights to the program, so that it can run on the remote machine. The problem in this is that the user has very little control over the remote machine, he has to delegate his rights to a program running there. In distributed systems, there is always a possibility that the remote machine is weakly protected and a malicious user can exploit the user's rights.

The authentication in distributed systems can be performed in two different ways. The decision is whether the security should be enforced centrally or locally.

- *In centralised* security enforcement, there could be some kind of Key Distribution Centre (KDC), where the keys of all the devices are stored. The Key Distribution Centre acts as a Trusted Third Party (TTP) that users can use to authenticate themselves and other users, and to get secure connections everywhere in the network. This solution expands the complexity, but it will also be more expensive to accomplish this kind of network. The biggest problem is the trustworthiness of the Trusted Third Party. If it is compromised, all the secret keys are available for malicious use and the whole scheme collapses.

- *In localised* security enforcement other kinds of security measures are needed. Each user enforces his own security policy, and trusts the machines he logs in to. There could be a trusted Certification Authority (CA), which issues public key certificates and a Certification Distribution Centre (CDC), which stores all the certificates issued by the Certification Authority. The users have their own key pairs and can certify their public keys with the Certification Authority. Then, if a user uses his key to sign something, the signature can be checked to correspond with a public key. The public key in turn can be checked with the CDC to certify that the public key in fact does belong to the user that originally did the signing. In this way, the security can be enforced locally and still have working authentication system with Public Key Infrastructure (PKI). [5]

2.3 Security in Ad Hoc Networks

In ad hoc networks there is no fixed infrastructure. All the devices are connected to each other via wireless links. Individual devices act as routers when relaying messages to other devices, which are too far apart from the sending one to get the message directly. The topology changes all the time in an ad hoc network when these mobile devices move in and out of other devices transmission range. All this makes the ad hoc networks very vulnerable to attacks and the security issues very complicated. [2]

2.3.1 Availability

Ensuring the availability is perhaps more important in ad hoc networks than it is in traditional networks. As all the devices in the network are dependent of each other to relay messages, and all the information is transmitted on the air, denial of service attacks is easy to perform. For example, a malicious user could try to jam or otherwise try to interfere with the flow of information on the air. Or it could be possible to disrupt the routing protocol used in the network by feeding the network with inaccurate information.

Routing protocols are in fact one of the most vulnerable points in ad hoc networks. They should be able to handle both the changing topology of the network and attacks from the malicious users. There are routing protocols that can adjust well to the changing topology, but all have problems in defending themselves against malicious attacks.

2.3.2 Authorisation and Key Management

Authorisation is another difficult matter in ad hoc networks. As there is very little or no infrastructure, so identifying users (e.g. participants in an ad hoc network in a meeting room) is difficult. There are some problems with trusted third party -schemes and identity-based mechanisms for key agreement. The trusted Certification Authority (CA) has to be available at all times. If the CA is unavailable, nodes cannot get the current public keys and are therefore unable to establish secure communication with others.

2.3.3 Confidentiality and Integrity

Confidentiality is also a vulnerable point in ad hoc networks. With wireless communication (as in WLAN), a user with proper equipment can sniff the messages on the air, and without proper encryption, the information will be available to anyone. On the other hand, without proper authentication, there is no point even to talk about confidentiality. If you cannot be certain who you are talking to, the confidentiality is poor anyway. And if the proper authenticity has been established, the securing of the connection with available keys is no problem.

For integrity, the same reasoning applies. In addition to malicious attacks, integrity may be compromised because of radio interference, etc., so some kind of integrity protection is definitely needed.

3 Bluetooth

In this section the project will give you generally understanding of Bluetooth and then go a bit more specific into the security solutions in Bluetooth.

Bluetooth is a low power short range radio technology, originally developed as a cable replacement to connect devices such as mobile phone handsets, headsets, and portable computers. Bluetooth has created a notion of a Personal Area Network (PAN), a close range wireless network that looks to revolutionise the way people interact with the information technology landscape around them. People no longer need to connect, plug into, install, enable or configure device to device for communications.
The Bluetooth specification is an open, global specification defining the complete system from the radio up to the application level. [7]

3.1 Background

Bluetooth started in 1994, when Ericsson Mobile Communications began a study to examine alternatives to the cables that linked its mobile phone to accessories.
The specification is named after Harald Blatand (Blatand is Danish for Bluetooth), Harald was the tenth-century Danish Viking king who united and controlled Denmark and Norway. The name was adopted because Bluetooth wireless technology is expected to unify the telecommunications and computing industries.
The Bluetooth Special Interest Group (SIG) was founded in February 1998, and is a group of companies working together to promote and define the Bluetooth specification. Initially the Bluetooth SIG consisted of the following group of core promoters; Ericsson, Nokia, Intel, IBM and Toshiba. Since then, almost all of the biggest companies in the telecommunications business (e.g. 3Com, Microsoft, Motorola) have joined the Bluetooth SIG and the number of the participating companies grew to 1,790 by April 2000. Version 1.0 of the Bluetooth specification was approved in the summer of 1999, and the latest version (at the time of writing) 1.1 in February 2001. [7]

Bluetooth can be used to connect almost any device to another device. The traditional example is to link a Personal Digital Assistant (PDA) or a laptop to a mobile phone. In that way you can easily make remote connections with your PDA or laptop without taking your mobile phone out from your pocket or messing around with cables. Bluetooth can also be used to form ad hoc networks of up to eight devices, called a piconet. This can be useful for example in a meeting, where all participants have their own Bluetooth-compatible laptops, and want to share files with each other.

3.2 Technical Specifications

Bluetooth devices are categorised into three different classes by the power they use. A class 1 device has a transmission power up to 100 mW (20 dBm) and a range up to 100 meters. A class 2 device has a transmission power of 1-2.5 mW (4 dBm) and a 10-meter range. A class 3 device has a 1 mW (0 dBm) transmission power and a range of 0.1-10 meters. There is also a minimum range for Bluetooth connection. If radios are put too close together, some receivers may saturate, but this is on short link lengths (below 10 cm). [7]

The architecture of Bluetooth is formed by the radio, the base frequency part and the Link Manager. Bluetooth uses the radio range around 2.4 GHz. The maximum bandwidth is 1 Mb/s, which is reduced by Forward Error Correction (FEC). Bluetooth specification designates the frequency hopping to be implemented with Gaussian Frequency Shift Keying (GFSK).

The base frequency part of the Bluetooth architecture uses a combination of circuit and packet switching technologies. Bluetooth can support either one asynchronous data channel and up to three simultaneous synchronous speech channels, or one channel that transfers asynchronous data and synchronous speech simultaneously.

The Link Manager is an essential part of the Bluetooth architecture. It uses Link Manager Protocol (LMP) to configure, authenticate and handle the connections between Bluetooth devices. It also operates the power management scheme, which is divided into three modes: sniff, hold and park.

3.2.1 The Bluetooth in relation to the OSI Reference Model

The most familiar reference model is probably the Open Systems Interconnect (OSI) standard reference model. The Bluetooth stack does not match the OSI-model exactly, but to comparison the different parts will be useful for understanding the different layers in the Bluetooth stack. Figure 1 show the relationship between the Bluetooth stack and the different layers in the OSI-model.

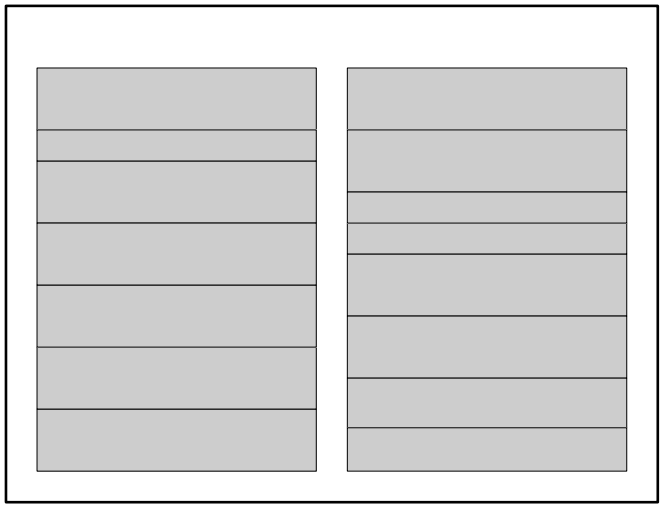

Figure 1 OSI reference model and Bluetooth

3.3 Bluetooth Protocol Architecture

This section describes some of the protocols in the Specification, their capabilities and the relation to each other. [1]

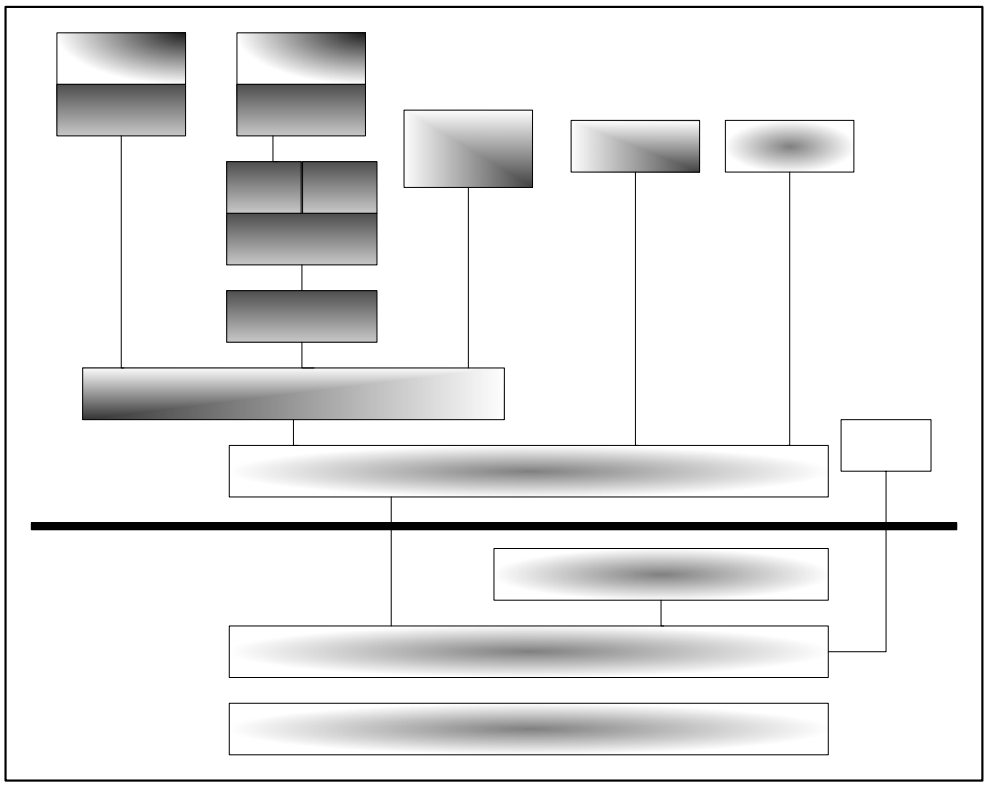

Figure 2 Bluetooth Protocol Stack

Figure 2 shows what the complete protocol stack comprises of both Bluetooth-specific protocols like LMP and L2CAP, and non-Bluetooth-specific protocols like OBEX (Object Exchange Protocol) and UDP (User Datagram Protocol).

It shows the relation how the protocols are using the services of other protocols when payload data needs to be transferred. The protocols may also have some other relations between the other protocols. For example some protocols (L2CAP, TCS Binary) may use LMP (Link Manager Protocol) when there is need to control the link manager.

vCard/vCal

OBEX

The Bluetooth protocol stack can be divided into four layers according to their purpose. This is shown in Table 1 The protocols and layers in the Bluetooth protocol stack.

Protocol layer	**Protocols in the stack**
Bluetooth Core Protocols	Baseband, LMP, L2CAP, SDP
Cable Replacement Protocol	RFCOMM
Telephony Control Protocols	TCS Binary, AT-commands
Adopted Protocols	PPP, UDP/TCP/IP, OBEX, WAP, vCard, vCal, IrMC1, WAE

Table 1 The protocols and layers in the Bluetooth protocol stack

In addition to the above protocol layers, the Specification also defines a Host Controller Interface (HCI), which provides a command interface to the baseband controller, link manager, and access to hardware status and control registers.

Together, the Cable Replacement layer, the Telephony Control layer, and the Adopted protocol layer form application-oriented protocols enabling applications to run over the Bluetooth Core protocols. The Bluetooth Specification is open and additional protocols (e.g., HTTP, FTP, etc.) can be accommodated in an interoperable fashion on top of the Bluetooth-specific transport protocols or on top of the application-oriented protocols.

3.3.1 Baseband

The Baseband and Link Control layer enables the physical RF link between Bluetooth units forming a piconet. As the Bluetooth RF system is a Frequency-Hopping-Spread-Spectrum system in which packets are transmitted in defined time slots on defined frequencies.

Baseband provides two different kind of physical links with their corresponding baseband packets, Synchronous Connection-Oriented (SCO) and Asynchronous Connectionless (ACL) which can be transmitted in a multiplexing manner on the same RF link. ACL packets are used for data only, while the SCO packet can contain audio only or a combination of audio and data. All audio and data packets can be provided with different levels of Forward Error Correction (FEC) or Cyclic Redundancy Check (CRC) error correction and can be encrypted. The different data types, including link management and control messages, are each allocated a special channel.

Compared to other wired physical media, the data packets defined by the Baseband Protocol are limited in size. Exporting a maximum transmission unit (MTU) associated with the largest Baseband payload (341 bytes for DH5 packets) limits the efficient use of bandwidth for higher layer protocols that are designed to use larger packets. Large L2CAP packets must be segmented into multiple smaller Baseband packets prior to their transmission over the air. Similarly, multiple received Baseband packets may be reassembled into a single larger L2CAP packet following a simple integrity check. The

Segmentation and Reassembly (SAR) functionality is absolutely necessary to support protocols using packets larger than those supported by the Baseband.

3.3.2 Link Manager Protocol

The link manager protocol is responsible for link set-up between Bluetooth devices. This includes security aspects like authentication and encryption by generating, exchanging and checking of link and encryption keys and the control and negotiation of baseband packet sizes.

3.3.3 Logical Link Control and Adaptation Protocol

The Bluetooth logical link control and adaptation protocol (L2CAP) adapts upper layer protocols over the baseband. It provides connection-oriented and connectionless data services to the upper layer protocols with protocol multiplexing capability, segmentation and reassembly operation. L2CAP permits higher level protocols and applications to transmit and receive L2CAP data packets up to 64 kilobytes in length.

3.3.4 PPP

In the Bluetooth technology, PPP is designed to run over RFCOMM to accomplish point-to-point connections. PPP is the IETF Point-to-Point Protocol and PPP-Networking is the means of taking IP packets to/from the PPP layer and placing them onto the LAN.

3.3.5 TCP/UDP/IP

These protocol standards are defined by the Internet Engineering Task Force and used for communication across the Internet. The implementation of these standards in Bluetooth devices allows for communication with any other device connected to the Internet, e.g. a Bluetooth cellular handset or a data access point is then used as a bridge to the Internet. UDP/IP/PPP is also available to transport WAP.

3.3.6 WAP

The Wireless Application Protocol (WAP) Forum is building a wireless protocol specification [6] that works across a variety of wide-area wireless network technologies. The goal is to bring Internet content and telephony services to digital cellular phones and other wireless terminals. In Figure 3, the protocol stack of the WAP framework is shown.

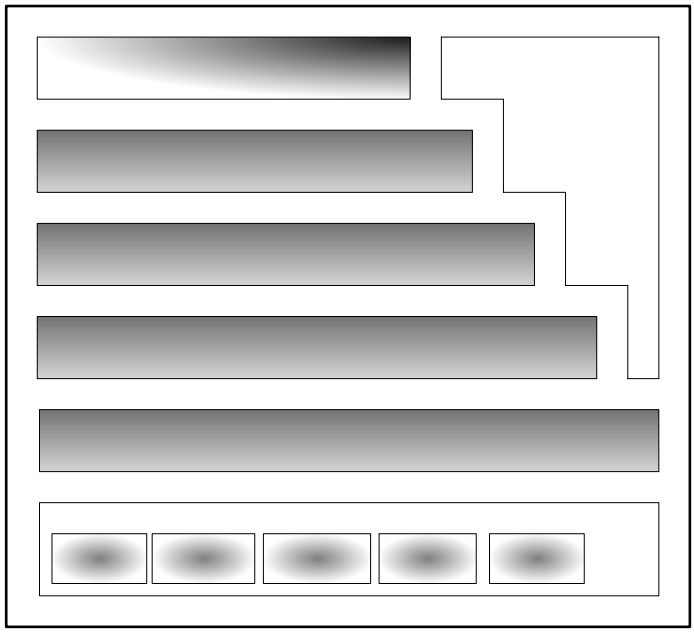

Figure 3 WAP Framework

The idea behind the choice of WAP is to reuse the upper software applications developed for the WAP Application Environment (WAE). These include Wireless Markup Language (WML) and Wireless Telephony Applications (WTA) browsers that can interact with applications on the PC. Building application gateways which mediate between WAP servers and some other application on the PC makes it possible to implement various hidden computing functionality, like remote control, data fetching from PC to handset etc.

3.4 Bluetooth Security in general

In every Bluetooth device, there are four entities to maintain the security at the link level. The Bluetooth device address (BD_ADDR), which is a 48-bit address that is unique to each Bluetooth device. The private authentication key, which is a 128-bit random number, is used for authentication purposes. The private encryption key, which may be between 8 and 128 bits in length, is used for encryption. And a random number (RAND), which is a frequently changing 128-bit random or pseudo-random number that is made by the Bluetooth device itself. [3, 12]

Generic Access Profile (GAP) defines the generic procedures related to discovery of Bluetooth devices (idle mode procedures) and link management aspects of connecting to Bluetooth devices (connecting mode procedures). It also defines procedures related to the use on different security levels. In addition, this profile includes common format requirements for parameters accessible on the user interface level.

Concerning security this is done in the Generic Access Profile. This profile specifies three security modes for a device:

- *Security Mode 1* (non-secure): A device will not initiate any security procedure. For example tasks that do not require security such as exchanging business cards.
- *Security Mode 2* (service level enforced security): A device does not initiate security procedures before channel establishment at L2CAP level. This mode allows different and flexible access policies for applications, especially running applications with different security requirements in parallel.
- *Security Mode 3* (link level enforced security): A device initiates security procedures before the link set-up at the Link Manager Protocol (LMP) level is completed.

Without regard to which level of security that is used, none of them prevent the optional use of higher layer security mechanisms. This may be VPN, IPSEC, TLS/WTLS, application level security, etc.

3.4.1 Security Levels

First the definition between authentication and authorisation:

- *Authentication* is the process of verifying "who" is at the other end of the link. In Bluetooth this is performed for devices (BD_ADDR). It is achieved by the authentication procedure based on the stored link key, which is a secret key shared between the master and slave, or by pairing (entering a PIN).
- *Authorisation* is the process of deciding if device X is allowed to have access to service Y. This is where the concept of 'trusted' exists. Trusted devices (authenticated and indicated as "trusted"), are allowed access to services. Untrusted or unknown devices may require authorisation based on user interaction before access to services is granted. This does not principally exclude that the authorisation might be given by an application automatically. Authorisation always includes authentication.

It is possible to define different security levels for devices and services.

For devices two trust levels are distinguished:

- *Trusted Device:* Device with fixed relationship (paired) that is trusted and has unrestricted access to all services. This device has been previously authenticated, a link key is stored and the device is marked as "trusted" in the Device Database.
- *Untrusted Device:* Device with no permanent fixed relationship (but possibly a temporary one) or device that has a fixed relationship, but is not considered as trusted. The access to services is restricted. This device has been previously authenticated, a link key is stored but the device is not marked as "trusted" in the Device Database.

- *Unknown Device:* No security information is available for this device. This is also an untrusted device.

For services the requirement for authentication, authorisation and encryption are set independently. The access requirements allow three security levels to be defined:
- *Services that require authentication and authorisation.* Automatic access is only granted to trusted devices. Other devices need manual authorisation.
- *Services that require authentication only.* Authorisation is not necessary.
- *Services open to all devices*; authentication is not required, no access approval required before service access is granted.

But there is flexibility between these security levels. It is possible to grant access to some services without providing access to other services (example: On a cellular phone, Service Discovery records shall be accessible, whereas dialup networking shall only be available for specific devices). The security architecture supports security policies for devices with some services communicating with changing remote devices (example: File Bluetooth Security Transfer or Business Card Exchange). Access granted to a service on such a device does not open up access to other services on the device, does not grant future access automatically or in an uncontrolled way to services on the device.

About implementation and Bluetooth-specific matters:
- The security architecture allows different protocols to enforce the security policies. For example, L2CAP will enforce the Bluetooth security policy for cordless telephony, RFCOMM will enforce the Bluetooth security policy for dialup networking, and OBEX (Object Exchange Protocol) will use its own security policy for file transfer and synchronisation.
- The architecture can completely work using security mode 2 of the Generic Access Profile. Especially since there are no changes to Baseband and LMP functions for authentication and encryption.
- The enforcement policy for authentication, authorisation or encryption might be different for client and server role. The security level of peer entities running an application needs is not required be symmetric.

3.4.2 Key Management

All security transactions between two or more parties are handled by the link key. The link key is a 128-bit random number. It is used in the authentication process and as a parameter when deriving the encryption key. The lifetime of a link key depends on whether it is a semi-permanent or a temporary key. A semi-permanent key can be used after the current session is over, to authenticate Bluetooth units that share it. A temporary key lasts only until the current session is terminated and it cannot be reused. Temporary keys are commonly used in point-to-multipoint connections, where the same information is transmitted to several recipients.

There are several different types of keys defined in Bluetooth. Link keys can be; combination keys, unit keys, master keys or initialisation keys, depending on the type of application. In addition to link keys, there is the encryption key.

The unit key is generated in a single device when it is installed. The combination key is derived from information from two devices and it is generated for each new pair of Bluetooth devices. The master key is a temporary key, which replaces the current link key. It can be used when the master unit wants to transmit information to more than one recipient. The initialisation key is used as a link key during the initialisation process when there are not yet any unit or combination keys. It is used only during the installation.

The length of the Personal Identification Number (PIN) code used in Bluetooth devices can vary between 1 and 16 octets. The regular 4-digit code is sufficient for some applications, but higher security applications may need longer codes. The PIN code of the device can be fixed, so that it needs to be entered only to the device wishing to connect. Another possibility is that the PIN code must be entered to the both devices during the initialisation.

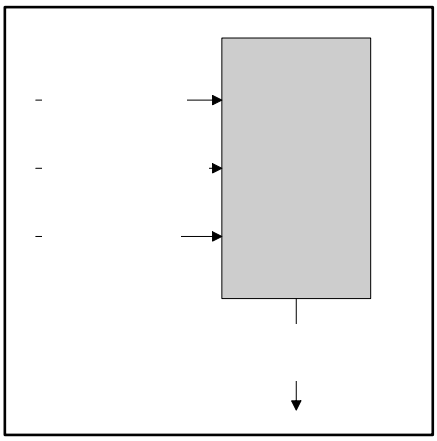

Figure 4 Key generating algorithm E22 for master and initialisation keys

The initialisation key is needed when two devices with no prior communications need to communicate. The key is used to protect the transfer of initialisation parameters. During the initialisation process, the PIN code is entered in both devices. The initialisation key is generated by the E22 algorithm, which uses the PIN code, the Bluetooth Device Address (BD_ADDR) of the claimant device and a random number generated by the verifier device as its parameters. The resulting 128-bit initialisation key is used for key exchange during the generation of a link key. After the key exchange the initialisation key is discarded.

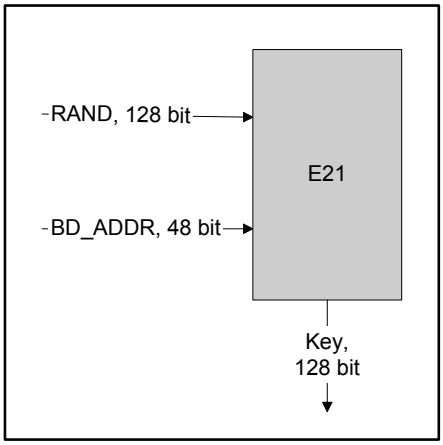

Figure 5 Key generating algorithm E21 for unit and combination keys

The unit key is generated with the key generating algorithm E21 when the Bluetooth device is in operation for the first time. Another device can use the other devices unit key as a link key between these devices. During the initialisation process, the application decides which party should provide its unit key as the link key. If one of the devices is of restricted memory capabilities (i.e. cannot remember any extra keys), its link key is to be used.

The combination key is generated during the initialisation process, and a new one is generated for each new combination of two Bluetooth units. It is generated by both devices at the same time, derived from information in both units. First, both of the units generate a random number. With the key generating algorithm E21, both devices generate a key, combining the random number and their Bluetooth device address. Thereafter, the devices securely exchange their random numbers and calculate the combination key to be used between them.

The master key is the only temporary key of the link keys described above. It is generated by the master device by using the key generating algorithm E22 with two 128-bit random numbers. The reason for using the key generating algorithm in the first place is to make sure the resulting random number is random enough. A third random number is then transmitted to the slave and with the key generating algorithm and the current link key an overlay is computed by both the master and the slave. The new link key (the master key) is then sent to the slave and bitwise XORed with the overlay. With this, the slave can calculate the master key. This procedure must be performed with each slave the master wants to use the master key with.

The encryption key is generated with the key generating algorithm E3 from the current link key, a 96-bit Ciphering Offset Number (COF) and a 128-bit random number. The COF is based on the Authenticated Ciphering Offset (ACO), which is generated during the authentication process. When the Link Manager (LM) activates the encryption, the

encryption key is generated. It is automatically changed every time the Bluetooth device enters the encryption mode.

3.4.3 Authentication

The Bluetooth authentication scheme uses a challenge-response strategy, where a 2-move protocol is used to check whether the other party knows the secret key. The protocol uses symmetric keys, so a successful authentication is based on the fact that both participants share the same key. As a side product, the ACO is computed and stored in both devices and is used for cipher key generation later on.

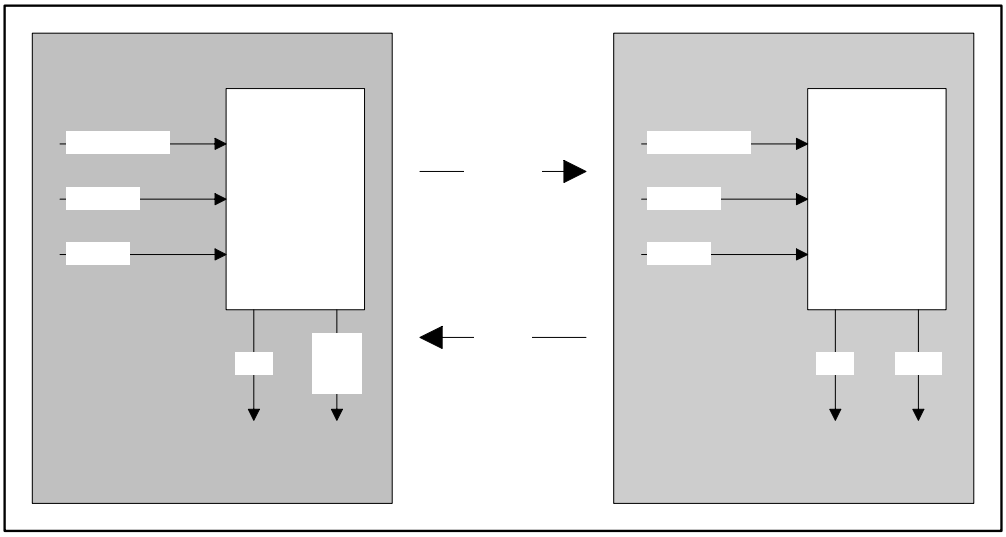

Figure 6 The authentication process

First, the verifier sends the claimant a random number to be authenticated. Then, both participants use the authentication function E1 with the random number, the claimants Bluetooth Device Address and the current link key to get a response. The claimant sends the response to the verifier, who then makes sure the responses match.

The application being used indicates who is to be authenticated. So the verifier does not have to be the master. Some applications require only one way authentication, so that only one party is authenticated, and others require mutual authentication, where both parties have to be authenticated in turn.

If the authentication fails, there is a period of time that must pass until a new attempt at authentication can be made. The period of time doubles for each subsequent failed attempt from the same address, until the maximum waiting time is reached. The waiting time decreases exponentially to a minimum when no failed authentication attempts are made during a time period. The waiting interval depends on the implementation.

3.4.4 Encryption

The Bluetooth encryption system encrypts the payloads of the packets. This is done with a stream cipher E0, which is re-synchronised for every payload. The E0 stream cipher consists of the payload key generator, the key stream generator and the encryption/decryption part.

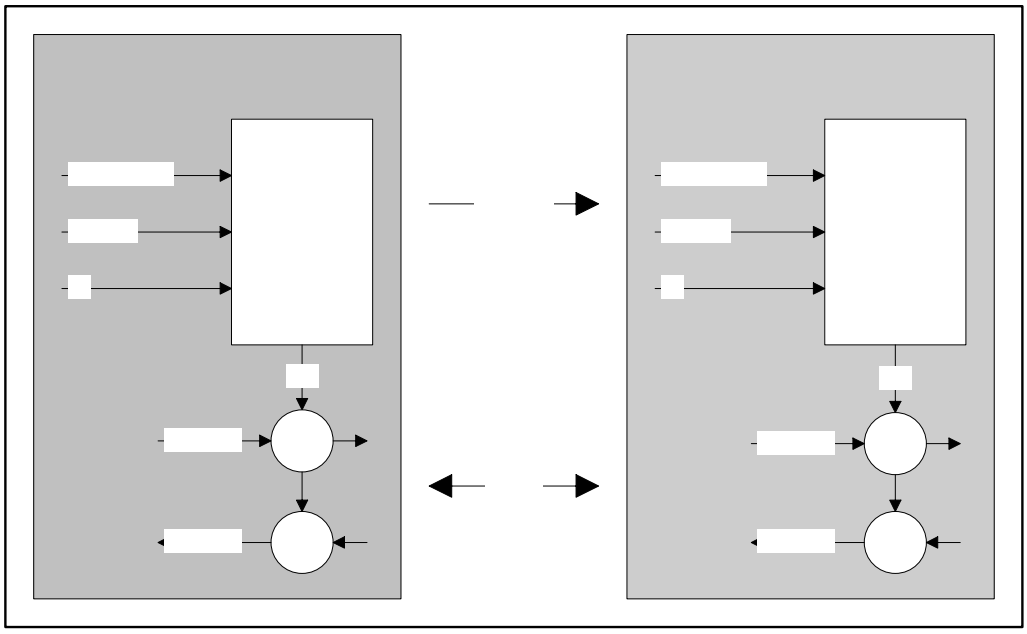

Figure 7 The encryption process

The payload key generator combines the input bits in an appropriate order and shifts them to the four Linear Feedback Shift Registers (LFSR) of the key stream generator. The key stream bits are generated by a method derived from the summation stream cipher generator attributable to Massey and Rueppel. The method has been thoroughly investigated, and there exist good estimates of its strength with respect to presently known methods for cryptanalysis. Although the summation generator has weaknesses that can be used in so-called correlation attacks, the high re-synchronization frequency will disrupt such attacks.

Depending on whether a device uses a semi-permanent link key or a master key, there are several encryption modes available:

- *If a unit key or a combination key is used*, broadcast traffic is not encrypted. Individually addressed traffic can be either encrypted or not.
- *If a master key is used*, there are three possible modes.
    - *Encryption mode 1*, nothing is encrypted.

- *Encryption mode 2*, broadcast traffic is not encrypted, but individually addressed traffic is encrypted with the master key.
- *Encryption mode 3*, all traffic is encrypted with the master key.

As the encryption key size varies from 8 bits to 128 bits, the size of the encryption key used between two devices must be negotiated. In each device, there is a parameter defining the maximum allowed key length. In the key size negotiation, the master sends its suggestion for the encryption key size to the slave. The slave can either accept and acknowledge it, or send another suggestion. This is continued, until a consensus is reached or one of the devices aborts the negotiation. The abortion of the negotiation is done by the application. In every application, there is defined a minimum acceptable key size, and if the requirement is not met by either of the participants, the application aborts the negotiation and the encryption cannot be used. This is necessary to avoid the situation where a malicious device forces the encryption to be low in order to do some harm.

For link encryption and authentication, Bluetooth uses a contemporary cipher algorithm called SAFER+ (Secure And Fast Encryption Routine), which generates 128-bit cipher keys from a 128-bit plain text input. [7]

In Bluetooth, the plaintext is provided by a combination of a predefined device PIN number or a unit key and random number. The resulting key is then loaded together with the BD (Bluetooth Devise) address, Master clock bits, and another 128 bit random number into a bank of Linear Feedback Shift Registers (LFSRs). The output of these LFSRs is combined by a Finite State Machine (FSM) called the “Summation Combiner” to produce a cipher stream which is then exclusive-OR’d (XOR’d) with either the transmit or receive data streams as required.
The LFSR block and Summation Combiner are together referred at the “Encryption Engine” and this process as the “E0” algorithm. This is the part that actually encrypts or decrypts the data bitstream, while the key generator is the part that uses the SAFER+ algorithm to generate the keys used by E0.

3.4.5 Ad Hoc Aspects

There are some aspects of Bluetooth security that should be considered in the light of ad hoc networking. In an ad hoc network formed in a conference room, there are a couple of possibilities for Bluetooth devices to secure the traffic. First of all, they can use the combination keys to encrypt the traffic. This means that the master device forms combination keys with every slave device in the network. Then the information from a slave is subsequently sent to all other slaves by the master.

Another way of forming a secure ad hoc network is to use the master key concept. Then all the devices in the network can use the same key when encrypting the traffic and no separate relaying of traffic is needed.

This seems to be the limit of the ad hoc aspects of the link level security mechanisms of Bluetooth. If there is to be more complex ad hoc networking, the security must be done

on the application level. For example, if any Key Distribution Centres (KDCs) or distributed secret schemes have to be used, Bluetooth does not support them directly.

3.5 Evaluation of the Security in Bluetooth

There is a problem in the usability of the Bluetooth devices. The use of the PIN code in the initialisation process of two Bluetooth devices is tacky. When you have to enter the PIN code twice every time you connect two devices, it gets annoying even with shorter codes. If there is an ad hoc network of Bluetooth devices and every machine is to be initialised separately, it is unbearable. And it does not make upholding the security very easy.

The generation of the initialisation key may also be of concern. The strength of the initialisation key is based on the PIN code used. The E22 initialisation key generation algorithm derives the key from the PIN code, the length of the PIN code and a random number, which is transmitted over the air. The output is some kind of questionable, as the only secret is the PIN code.

The problem of the usability and the strength of the initialisation key may be solved with application level key agreement. The specification makes a suggestion to use application level key agreement software with longer (between 1 and 16 octets) PIN codes. So the PIN code does not need to be entered physically to each device of the connection, but is exchanged by, for example Diffie-Hellman key agreement.

There is another problem in the unit key scheme. Authentication and encryption are based on the assumption that the link key is the participants' shared secret. All other information used in the procedures is public. Now, suppose that devices A and B use A's unit key as their link key. At the same time (or later on), device C may communicate with device A and use A's unit key as the link key. This means that device B, having obtained A's unit key earlier, can use the unit key with a faked device address to calculate the encryption key and therefore listen to the traffic. It can also authenticate itself to device A as device C and to device C as device A.
But there are some problems that have to be dealt with to perform this. In this case B has to know who A is communicating with, and fake the device address, which is in hardware and therefore hard to change. So there has to be some coincidence to perform this kind of attack successfully.

As mentioned earlier, it is possible to use application level key exchange and encryption methods to secure the communication, on top of the existing Bluetooth security systems. If certificate-based methods are employed, it is possible to defend against man-in-the-middle attacks.

4 GSM/GPRS

This chapter describes mainly the security that is provided by GPRS, but also some of the security in GSM which deal with the data communication. It starts with GPRS in general and ends with an evaluation of the security.

With the introduction of the General Packet Radio Service (GPRS) in existing GSM-networks, data services with high bandwidth can be provided to mobile users. [9]

4.1 GPRS in general

General Packet Radio Service (GPRS) provides short connection setup-times, virtual connections, and data rates up to 115 kbit/s for each user, while the available bandwidth can be shared among different users. The high bandwidth will be achieved by combining up to eight time slots at the radio interface, where the data is transported in a packet-oriented way.

While the radio interface of GPRS is very much the same like in GSM, a new core network is defined in parallel to the existing GSM network. As shown in Figure 8, the BSC splits the voice and data traffic. Voice traffic is sent to a traditional ISDN-based GSM-network while data traffic is transported via a separated IP-based backbone network.

The Base Transceiver Station (BTS) and the Base Station Controller (BSC) together form the access network of the GSM/GPRS network. The Mobile Switching Centre (MSC) is responsible for the routing of the calls, the tracking of the mobile users and security functions. The Visitor Location Register (VLR) is a database storing actual user related information of the users currently served by the MSC. The Home Location Register (HLR) holds further user information, like the actual location and the subscription data of the users.

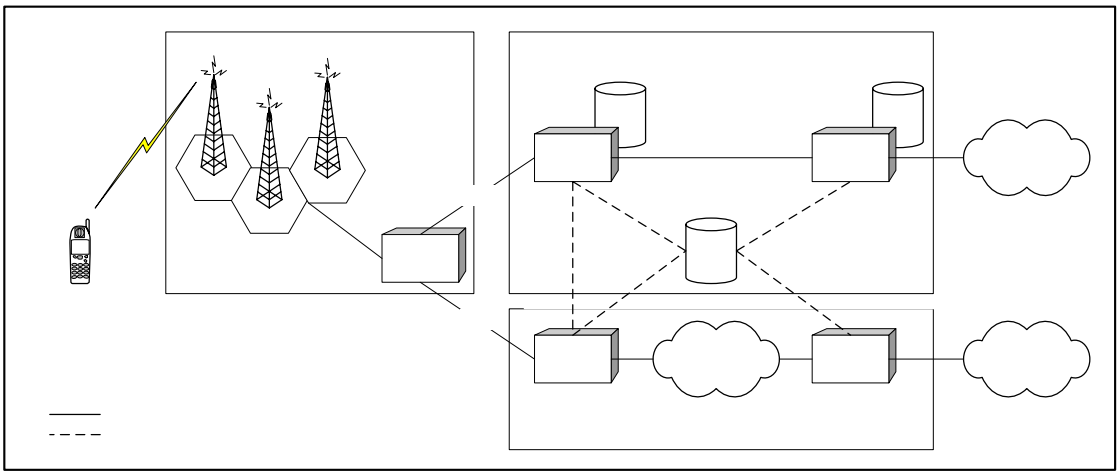

Figure 8 GSM/GPRS network architecture

The GSM network required some modifications to support General Packet Radio Service. GPRS attempts to reuse the existing GSM network elements as much as possible. But in order to effectively build a packet-based mobile cellular network, some new network elements, interfaces and protocols are required: [16]

- *The Subscriber Terminal (TE);* there is required a totally new GPRS terminal to get access to GPRS services, these terminals are also backward compatible with GSM.
- *Base Transceiver Station (BTS);* needs software upgrade.
- *Base Station Controller (BSC);* needs software upgrade, as well as installation of some extra hardware called Packet Control Unit (PCU), which directs the data traffic to the GPRS network.
- *Databases (VLR, HLR, etc);* all databases involved in the network needs software upgrades to handle the new call models and functions introduced by GPRS.
- *Serving GPRS Support Node (SGSN);* is one of the new nodes in the GPRS network architecture. It is at the same hierarchical level as the MSC, keeps track of the individual MS location and performs security functions and access control. The SGSN is connected to the base station system with Frame Relay.
- *Gateway GPRS Support Node (GGSN);* is the other node in the GPRS network architecture. It interacts with external packet-switched networks, and is connected to SGSN via an IP-based GPRS backbone network.

BTS

BTS

Three classes of GPRS MSs are supported:

- A *class-A MS* can operate GPRS and other GSM services simultaneously.
- A *class-B MS* can monitor control channels for GPRS and other GSM services simultaneously, but can only operate one set of services at one time.
- A *class-C* MS can exclusively operate GPRS services.

GSM / GPRS

In order to access the GPRS services, a MS shall first make its presence known to the network by performing a GPRS attach. This operation establishes a logical link between

Data/voice connection
SS7 based signaling

the MS and the SGSN, and makes the MS available for SMS over GPRS, paging via SGSN, and notification of incoming GPRS data.

To send and receive GPRS data, the MS has to activate the packet data address that it wants to use. This operation makes the MS known in the corresponding GGSN, and interacting with external data networks can begin.

The GPRS network encapsulates all data network protocols into its own encapsulation protocol, called the GPRS Tunnelling Protocol (GTP). This is done to simplify the routing mechanism and the delivery of data over the GPRS network.

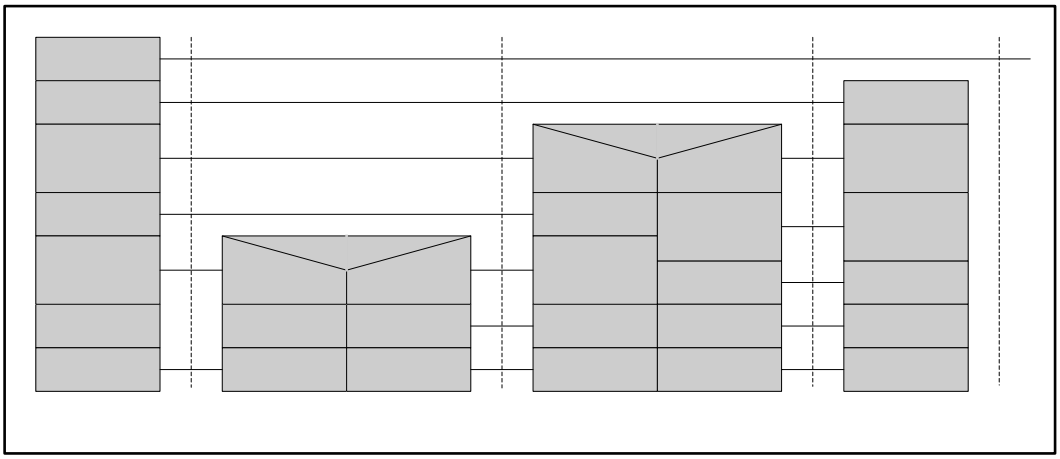

Figure 9 GPRS Network Protocol Stack

GTP packets carry the user's IP or X.25 packets. Below GTP, the standard protocols TCP or UDP are employed to transport the GTP packets within the backbone network. X.25 expects a reliable data link, thus TCP is used. UDP is used to access IP-based packet data networks, which do not expect reliability in the network layer or below. IP is employed in the network layer to route packets through the backbone. Ethernet, ISDN, or ATM-based protocols may be used below IP.

This transparent transfer method reduces the requirement for the GPRS Public Land Mobile Network (PLMN) to read external data protocols, and it enables easy introduction for additional protocols in the future.

4.2 GPRS Security

GPRS security is almost equivalent to the existing GSM security. The main entities involved are the Serving GPRS Support Node (SGSN), Gateway GPRS Support Node (GGSN), Authentication Centre (AuC), and Home Location Register (HLR). The HLR and the AuC provide the same functionality as in GSM. The SGSN performs authentication and cipher setting procedures based on the same algorithms, keys, and

criteria as in existing GSM. GPRS uses a ciphering algorithm optimised for packet data transmission.

The main security functions related to the GPRS device (MS) are authentication and encryption. Authentication is performed in the same way as in GSM with a challenge-response protocol. Encryption in GPRS differs a bit from encryption in GSM. A new A5 algorithm has been developed for GPRS, and encryption is done between the MS and the SGSN (instead of between the MS and the base station as in GSM).

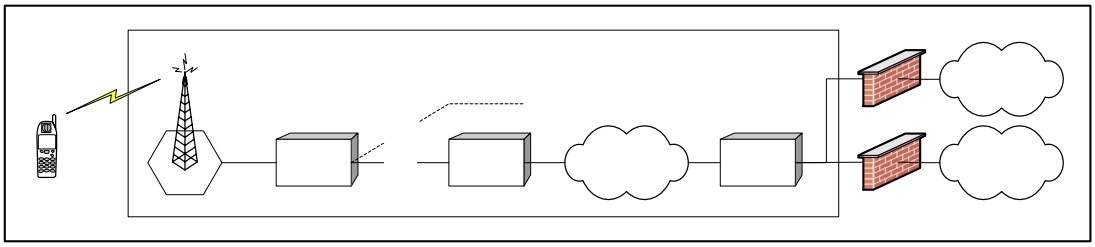

Figure 10 GPRS Encryption

GPRS backbone security is ensured by restricting access to the network, and by protecting the network boundaries by placing firewalls between the GPRS network and networks belonging to other operators as well as other external networks.

It is the operator that is responsible for the security of its own Intra-PLMN backbone which includes all network elements and physical connections. The operator shall prevent unauthorised access to its Intra-PLMN backbone. A secure Intra-PLMN backbone guarantees that no intruder can eavesdrop or modify user information and signalling in the Intra-PLMN backbone. [20]

The GPRS architecture utilises GPRS tunnelling and private IP addressing within the backbone to restrict unauthorised access. User traffic addressed to a network element shall be discarded. Firewall functionality may provide these services at the access points of the Intra-PLMN backbone.

The Inter-PLMN links shall be negotiated between operators as part of the roaming agreement. They shall ensure that the Inter-PLMN links are secure providing integrity and confidentiality. For example, secure links can be achieved by point to point links, private Inter-PLMN backbones or encrypted tunnels over the public Internet.

Operators shall be able to determine the origin of packets coming from the inter-PLMN backbone. One example is to use a Frame Relay PVC between two operators.

4.2.1 The authentication algorithm

The cellular phone (MS) consists of the mobile station itself (ME – Mobile Equipment) and a SIM-card (Subscriber Identity Module).

The mobile equipment consists of a:
- GPRS A5 algorithm to encrypt the data
- International Mobile Equipment Identity (IMEI) that is physically secured in the ME

The primary function of the SIM-card is to authenticate the MS before its get access to the network, and contains the:
- International Mobile Subscriber Identity (IMSI)
- Subscriber Identification Key (Ki)
- Authentication algorithm (A3)
- Ciphering key generating algorithm (A8)
- Personal Identification Number (PIN)

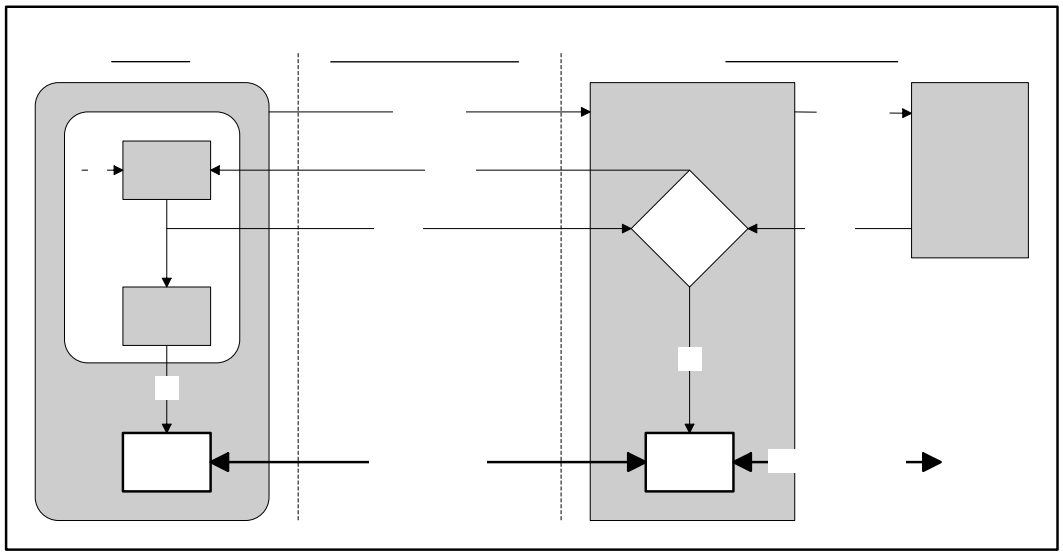

Figure 11 Authentication and key derivation

Authentication is performed by a challenge and response mechanism:
- The mobile sends an authentication request to the network. This arrives at the SGSN and is sent further to the HLR and AuC, which generates triplets. These triplets consist of a Random Number (RAND, 128bit), Signed Response (SRES, 32bit) and encryption key Kc (64bit). RAND and SRES are used as a challenge and response to authenticate the smart card in the mobile station.
- The triplets are sent to the SGSN, which sends a random challenge (RAND) to the mobile.
- The mobile encrypts the challenge using the authentication algorithm (A3) and the key assigned to the mobile (Ki), and sends a response back (SRES).

- This response is compared with the SRES in the authentication triplets. If they are identical, the MS must have the correct authentication algorithm and Ki, and are therefore genuine.
- The response is then passed through an algorithm A8 to derive the key Kc, used together with the frame number (22bit) to encrypt the signalling (114bit frames) between the MS and SGSN.

GPRS does not itself offer end-to-end security, but there are many proprietary and other standards that cover methods of end-to-end encryption, such as WAP forum or Internet security protocols covered by IPSec. Alternatively, organisations that set up Virtual Private Networks.

4.3 Security in GSM and GPRS

GSM and GPRS share authentication mechanisms and they both use encryption on the air interface: [19]

- GSM uses A5 encryption
- GPRS uses GEA (GEA2) encryption

By default, neither GSM nor GPRS provide any encryption beyond the air interface - there is no end-to-end encryption. Integrity is provided through encryption, which again means that there is no cryptographically valid end-to-end integrity protection.

4.3.1 Security in GSM

Confidentiality (Encryption - User data confidentiality)

- Only the network can initiate encryption (encryption is an option).
- The user receives no indication if encryption is provided or not.
- Encryption in GSM is comparably weak (although key is 64 bit, only 54 is used in practice, furthermore the national regulator and/or operator may weaken the encryption if they want).
- Encryption is prohibited is some countries and export restrictions have led to weaker encryption in other countries.
- Encryption is in principle only provided over the air interface - there is no end-to-end encryption provided by the system.

Confidentiality (Anonymity/User Identity Confidentiality)

- Primitive provision for anonymity of user movements by using the Temporary Mobile Subscriber Identity (TMSI).
- The network may at any time request the full identity (IMSI – International Mobile Subscriber Identity).

Integrity (of user data)

- User data integrity provided by encryption - which implies that there are no end-to-end integrity mechanisms.

Authentication (of User Identity)
- The network may authenticate the subscriber (usually does).
- The subscriber cannot authenticate the network.

Access control
- Access control is provided based on user identity authentication and subscription checking in HLR/VLR (the service list in HLR/VLR can be viewed as an Access Control List).

Denial-of-Service
- Very limited Denial-of-Service protection in GSM.

Non-Repudiation
- Very limited Non-Repudiation service in GSM although the event log/charging data can be viewed as a audit trail (but in general it does not suffice as a Secure Audit Trail).

4.3.2 Security in GPRS

Security in GPRS is almost identical to GSM in many respects:
- Same authentication mechanism
- Same access control mechanism
- Same crypto-key generating mechanism
- Same level of Denial-of-Service protection
- Same level of Non-Repudiation
- Same level of SIM card security
- Same level of confidentiality protection - although a new encryption algorithm (GEA/GEA2) to accommodate packet switching is introduced

Threats to the security in GPRS are very different from GSM:
- Circuit switched GSM does not provide much security, but then the security threats are also quite limited, because of the obscurity of the SS7 protocol.
- GPRS backbone is based on IP - GPRS will depend on sufficient IP security
    - The TCP/IP protocol suite is well known to a lot of people
    - It is almost trivial to automate large scale attacks on Internet nodes
    - The cost of mounting a large scale attack is negligible
- GGSN is external interface towards Internet
    - GGSN must be able to resist dedicated attacks

4.4 Vulnerabilities in Second Generation security

Second Generation involves both GSM and GPRS. 2G security got some weaknesses that 3G (UMTS) will improve but also elements that that will be retained. [18]

Weaknesses in Second Generation security:

- Active attacks using a "false BTS" are possible.
- Cipher keys and authentication data are transmitted in clear between and within networks.
- Encryption does not extend far enough towards the core network resulting in the cleartext transmission of user and signalling data across microwave links (in GSM, from the BTS to the BSC).
- User authentication using a previously generated cipher key (where user authentication using RAND, SRES and A3/8 is not provided) and the provision of protection against channel hijack rely on the use of encryption, which provides implicit user authentication. However, encryption is not used in some networks, leaving opportunities for fraud.
- Data integrity is not provided. Integrity defeats certain false BTS attacks and, in the absence of encryption, provides protection against channel hijack.
- The IMEI is an unsecured identity and should be treated as one.
- Second generation systems do not have the flexibility to upgrade and improve security functionality over time.

Second Generation Security Elements to be retained

Third Generation security shall retain, and in some cases develop, the following security elements of second generation systems:

- Authentication of subscribers for service access. Problems with inadequate algorithms will be addressed. Conditions regarding the optionally authentication and its relationship to encryption shall be clarified and tightened.
- Radio interface encryption. The strength of the encryption will be greater than that used in second generation systems (the strength is a combination of key length and algorithm design). This is to meet the threat posed by the increased computing power available to those attempting cryptanalysis of the radio interface encryption.
- Subscriber identity confidentiality on the radio interface. However, a more secure mechanism will be provided.
- The SIM as a removable, hardware security module that is:
    - Manageable by network operators.
    - Independent of the terminal as regards its security functionality.
- SIM application toolkit security features providing a secure application layer channel between the SIM and a home network server. Other application layer channels may also be provided.
- The operation of security features is independent of the user, i.e. the user does not have to do anything for the security features to be in operation. However, greater user visibility of the operation of security features will be provided to the user.

5 IPSec

IPSec is an Internet Engineering Task Force (IETF) standard suite of protocols that provides data authentication, integrity, and confidentiality as data is transferred between communication points across IP networks. IPSec is a security addition to the IP protocol, which enables security and privacy to TCP/IP communication, and provides data security at the IP packet level. [10]

To ensure privacy, data is encrypted with an encryption algorithm called triple DES (3DES), and are one of the widest used algorithms for strong encryption. Triple DES has an effective number of keys that is approximately 2^112. If you feel lucky and think that you only have to try half of the keys to find the right one, you would still have to try 2^111 keys. To get an idea of how big the number 2^112 is, the odds of being killed by a lightning is 2^33 per day.

Normally IPSec consists of two parts, the key management (IKE – Internet Key Exchange) and the encryption (ESP – Encapsulating Security Protocol). It is the most widely used protocol for VPN's (Virtual Private Networks).

IPSec contains the following protocols:
- *Encapsulating Security Payload (ESP):* Provides confidentiality, authentication, and integrity.
- *Authentication Header (AH):* Provides authentication and integrity.
- *Internet Key Exchange (IKE):* Provides key management and Security Association (SA) management.

5.1 Encapsulating Security Payload (ESP)

ESP provides authentication, integrity, and confidentiality.

IPSec provides an open framework for implementing industry standard algorithms, such as SHA and MD5. The algorithms IPSec uses produce an identifier for each packet, which is equivalent to a fingerprint. This identifier allows the device to determine if a packet has been changed. Furthermore, packets that are not authenticated are discarded and not delivered to the intended receiver.

ESP also provides all encryption services in IPSec. Encryption translates the message into an unreadable format to hide the message content. This allows only the sender and the authorised receiver to read the data. In addition, ESP has an option to perform authentication, called ESP authentication. Using this, the ESP provides authentication and integrity for the payload but not for the IP header.

The ESP header is inserted into the packet between the IP header and any subsequent packet contents. However, because ESP encrypts the data, the payload is changed. ESP does not encrypt the ESP header, or the ESP authentication.

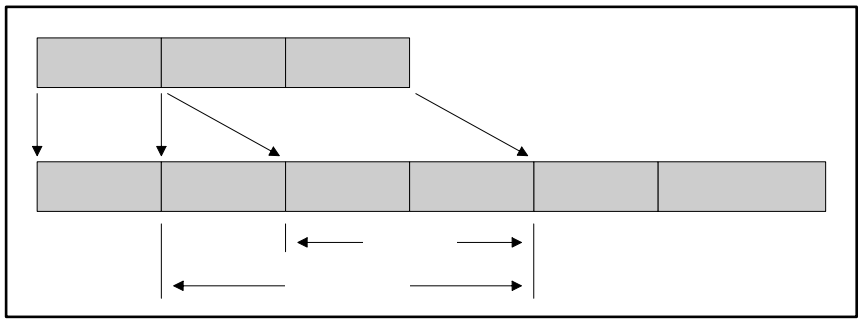

Figure 12 Packet with IPSec Encapsulating Security Payload (ESP)

5.2 Authentication Header (AH)

AH provides authentication and integrity. It also provides optional anti-replay protection, which protects against unauthorised retransmission of packets. The authentication header is inserted into the packet between the IP header and any subsequent packet contents. The payload is not touched. Although AH protects the packets origin, destination, and contents from being changed, the identity of the sender and receiver is known. The AH does not protect the confidentiality of the data. If data is intercepted and only AH is used, the message contents can be read.

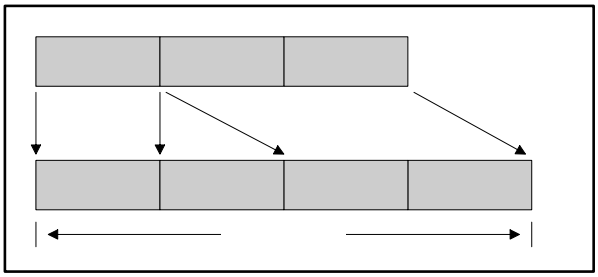

Figure 13 Packet with IPSec - AH - Transport Mode

IP HDR

IP HDR

5.3 Security Association (SA)

IPSec introduces the concept of the Security Association (SA). An SA is a logical connection between two devices transferring data. It provides data protection for unidirectional traffic by using the defined IPSec protocols. It is uniquely identified by a triple consisting of a Security Parameter Index (SPI), an IP Destination Address, and a security protocol (AH or ESP) identifier. An IPSec tunnel consists normally of two unidirectional SAs, which together provide a protected, full-duplex data channel.

Security Association Modes

SAs operate using modes. A mode is the method in which the IPSec protocol is applied to the packet. IPSec can be used in tunnel mode or transport mode. Typically, the tunnel mode is used for gateway-to-gateway IPSec tunnel protection, while transport mode is used for host-to-host IPSec tunnel protection.

- *Transport Mode* encapsulates only the packet's payload, the IP header is not changed. After the packet is processed with IPSec, the new IP packet contains the old IP header and the processed packet payload. Transport mode does not shield the information in the IP header, so an attacker can read where the packet is coming from and where it is going to. This mode can only be used when both the source and the destination system understand IPSec. The advantage of transport mode is that it adds only a few bytes to each packet, and allows implementation of IPSec in the network architecture without modifying the operating system or any applications on PCs, servers or hosts.
- *Tunnel Mode* encapsulates the entire IP packet, it becomes the payload of the packet that is processed with IPSec. A new IP header is created that contains the two IPSec gateway addresses. The gateways perform the encapsulation/decapsulation on behalf of the hosts. Tunnel mode ESP prevents an attacker from analysing the data and deciphering it, as well as knowing who the packet is from and where it is going to.

AH and ESP can be used in both transport mode and tunnel mode.

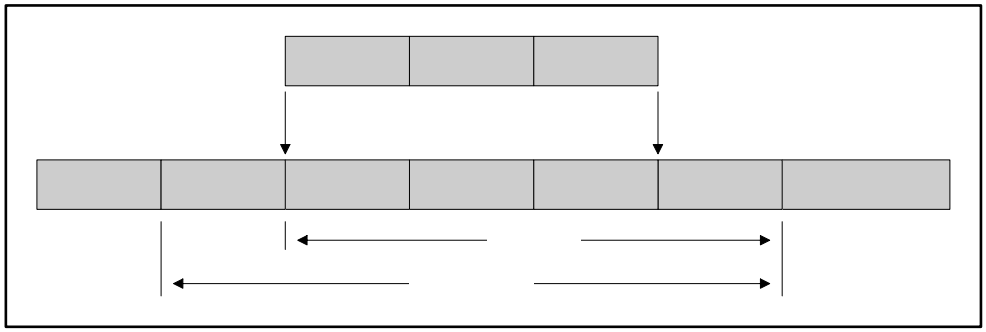

Figure 14 Packet with IPSec - ESP - Tunnel Mode

5.4 Key Management

IPSec uses the Internet Key Exchange (IKE) protocol to facilitate and automate the SA setup and the exchange of keys between parties transferring data. Using keys ensures that only the sender and receiver of a message can access the information.

IPSec requires that keys can be re-created, or refreshed, frequently, so that the parties can communicate securely with each other. IKE manages the process of refreshing keys, a user can control the key strength and the refresh frequency. Refreshing keys on a regular basis ensures data confidentiality between sender and receiver.

Internet Key Exchange (IKE) Setup
IKE works in a two-phase process. The first phase sets up the actual IKE SAs. The second phase sets up the secure data transmission channels, which are the IPSec SAs.

The first phase includes these tasks:
1. Two parties negotiate the encryption and authentication algorithms to use in the IKE SAs.
2. Then they authenticate each other using a predetermined mechanism, such as pre-shared keys or digital certificates.
3. A shared master key is generated by the Diffie-Hellman Public key algorithm within the IKE framework for the two parties. The master key is also used in the second phase to derive IPSec keys for the SAs.

The second phase includes these tasks:
1. The two parties negotiate the encryption and authentication algorithms to use in the IPSec SAs.
2. The master key is used to derive the IPSec keys for the SAs. Once the SA keys are created and exchanged, the IPSec SAs are ready to protect user data between the two VPN gateways.

5.5 Weaknesses in IPSec

Regarding Encrypted Data

The primary purpose of encryption is privacy. An attacker who can read other people's messages has completely defeated the security system. [21]

Stream ciphers such as RC4 have a serious disadvantage; Changes to the cipher text show up as predictable changes to the decrypted plaintext. Suppose that an attacker can trick a machine to send a known message to a target. This package can be intercepted, modified, and reinserted into the communication.

Defences

The proper defence against many attacks is use of integrity-checking. If a message is properly checked, it cannot be cut apart and modified to be sent away again. More precisely, all received messages should be checked for integrity, using acceptably strong cryptographic techniques. The problem is that even ESP or AH got a separate mechanism for integrity, however, authentication protect the integrity of the message.

A second defence technique is to avoid reuse of keying material for more than one connection. An attacker cannot cut and paste between connections if they use different keys, the inserted material will not be decrypted properly.
If this is not feasible, keys should be changed frequently. For stream ciphers it is important that this is based on time, data received, and too large difference in the indicated sequence number.

6 Wireless Application Protocol – WAP

In this section the project will show you different aspects of the Wireless Application Protocol. First generally about WAP and how it works, then more specific on the security mechanisms that WAP may offer.

WAP is a wireless protocol that allows mobile devices to use data service, applications and access the Internet. Services and applications may be e-mail, weather and traffic alerts, news, e-commerce, banking services, online address book, etc. WAP is able to work with a variety of different wireless technologies, each of which connects at the bottom of the WAP stack as a bearer. [7]

WAP uses a combination of Internet protocols (such as UDP) and protocols specially modified to work with mobile devices (e.g. WML). Figure 15 WAP on the Bluetooth protocol stack in chapter 6.1) shows how WAP lies above the Bluetooth in the protocol stack.

WAP uses a client/server architecture. Normally a WAP client will be a mobile device with lower bandwidth links than others that are wired to the network. Because of their low bandwidth links, WAP content is communicated in a compact format. WAP supports Wireless Markup Language (WML), which is similar to the Hyper-text Markup Language (HTML). Both are derived from a set of rules for producing markup languages called Extensible Markup Language (XML).

WAP defines a set of protocols in transport, security, transaction, session, and application layers to enable a creation of advanced mobile services. WAP is independent of the bearer services.

WAP is developed by WAP Forum industry association to provide specifications for the applications that operate over wireless communication networks. [6]

Internet technology could not be directly adopted because of the limitations of mobile terminals; they have less powerful CPUs and memory, restricted power consumption, smaller displays and different input devices. The mobile network has also some limitations that must be taken into account. These include less bandwidth, more latency, less connection stability and less predictable availability.

6.1 WAP over Bluetooth

WAP was designed for thin-client devices, such as mobile phones. Bluetooth simply provides another possible bearer beneath the WAP stack.
WAP over Bluetooth is defined in the Bluetooth standard.

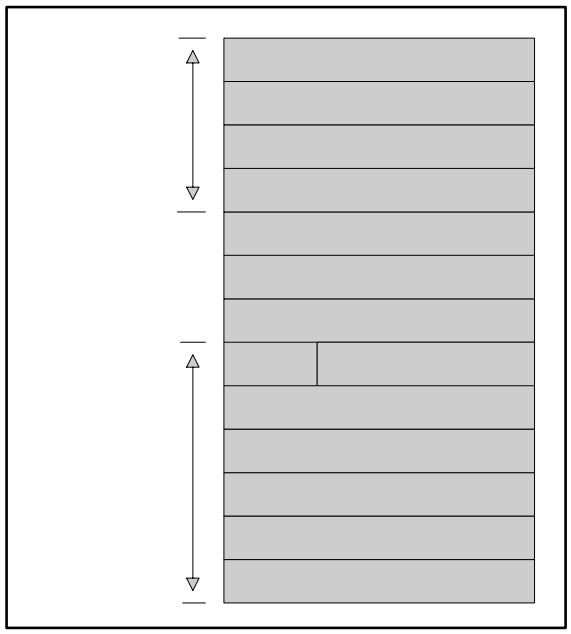

Figure 15 WAP on the Bluetooth protocol stack

WAE (Wireless Application Environment) – Provides a user interface, typically a micro browser, which is a lightweight version of a Web browser.
WSP (Wireless Session Protocol) – The WSP provides ways to establish a session from client to server, agree on used protocol functionality, exchange content, and suspend and resume sessions.
WTP (Wireless Transport Protocol) – Provides a reliable transport layer for WSP. But since the Bluetooth baseband provides reliable transport, WTP could be omitted when Bluetooth is used as a bearer for WAP.
WTLS (Wireless Transport Layer Security) – Provides security. This layer can be omitted for applications that do not require more security than Bluetooth baseband provides.
UDP (User Datagram Protocol) – Provides unreliable, connectionless datagram transport.
IP (Internet Protocol) – A protocol that supports addressing, routing, segmentation and reassembly of packets.
PPP (Point to Point Protocol) – A client/server-based packet transport system commonly used on dial-up links to carry IP traffic.

6.1.1 WAP over Bluetooth Applications

An application could use WAP over a Bluetooth phone as a two way, interactive Remote Control. The user controls a Bluetooth device by browsing the device's WAP pages,

which contain special links that trigger the device's functions. Examples would be to use the phone as light switch or door key, or to control the home alarm system. Let's take the Bluetooth door lock as an example: Once the user approaches a door, he actively connects to, or gets connected by the door and the door's WAP homepage is automatically displayed allowing the user to select an action (lock/ unlock). This action is basically a WAP request sent to the WAP server and the server executes a script triggering the actual lock or unlock operation. [15]

6.2 WAP over GPRS

When WAP content is delivered over a packet-switched service such as GPRS, subscribers can have instant access to WAP services.

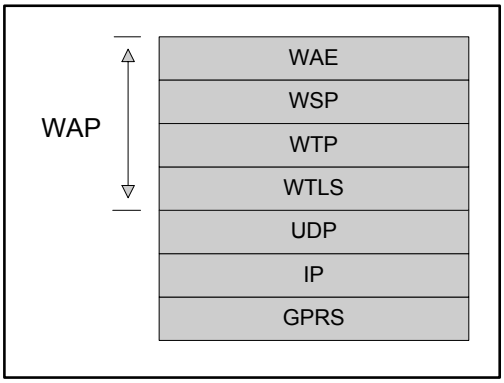

Figure 16 WAP on the GPRS protocol stack

Security in WAP and GPRS is performed at different protocol layers. GPRS provides security at the bearer layer, while WAP adds security on top of the transport layer.

GPRS provides a high level of security for all applications. The fist step is to authenticate the mobile terminal with standardised GSM techniques. After that, all data between the mobile terminal and the Serving GPRS Support Node (SGSN) is encrypted.

WAP security is based on Wireless Transport Layer Security (WTLS), which is independent of the underlying bearer. For establishing secure connections between the WAP gateway and the content servers on the Internet, Secure Sockets Layer (SSL) is used.

6.3 Security in the WTLS

A huge growth of the wireless mobile services poses demand for the end-to-end secure connections. [8] The Wireless Transport Layer Security (WTLS) protocol is the first attempt to provide a secure end-to-end connection for the Wireless Application Protocol. WTLS is based on the most common protocols, such as the widely used Transport Layer Security (TLS) protocol version 1.0 and the Secure Socket Layer (SSL) protocol version 3.0. However, it was not possible to apply the procedures used in the traditional connection-oriented world. The development work resulted in a protocol that resembles the TLS but it has some properties in order to adjust to the wireless world.

The Wireless Transport Layer Security protocol is the security layer of the WAP. Its primary goal is to provide privacy, data integrity, and authentication for WAP applications. Both the client and the server have to be authenticated and the connection must be encrypted. Man-in-the-middle attacks should be prevented so that the data cannot be modified during the transfer. The subscriber wants to be sure that the service being used is really the one it claims to be. Although the traffic in the air is encrypted in several mobile networks, however, the mobile network does not provide the complete end-to-end security. That is the reason why the WTLS is needed.

WTLS provides the upper-level layer for WAP with a secure transport service interface that preserves the transport service interface below it. It provides an interface for managing secure connections.

The wireless networks require support for both datagram and connection oriented transport layer protocols. The mobile equipment sets requirements for the algorithms because of the limited available processing power and memory. WTLS has been optimised for low-bandwidth bearer networks with relatively long latency. Fast algorithms are chosen into the algorithm.

6.3.1 Wireless Transport Layer Security

Security in the WAP architecture should enable services to be extended over mobile networks while also preserving the integrity of user data. Denial of service should also be prevented. The wireless mobile networks set many provisions in the security layer. [6]

One of the most important of the requirements is to support low data transfer rates. The amount of overhead must be kept as small as possible because of the low bandwidth. Other issues include slow interactions, limited processing power and memory capacity. Which also include the restrictions on exporting and using cryptography. The round-trip times can be long and the connection should not be closed because of that. The crypto-graphical algorithms must be light enough so that the mobile terminals are able to execute them. The numbers of crypto-graphical algorithms has to be minimised and small-sized algorithms must be used, because of the small amount of RAM in the mobile terminals.

In short, the objective of the WTLS is to be a lightweight and efficient protocol with respect to bandwidth, memory and processing power.

Specification
The WTLS layer operates above the transport protocol layer and it provides the upper level layer of the WAP with a secure transport service interface. It also presents methods to manage secure connections.

The WAP, by means of the WTLS, provides end-to-end security between the WAP protocol endpoints, the end points are the mobile terminal and the WAP gateway. When the WAP gateway makes the request to the origin server, it will use the SSL below HTTP to secure the request. This means that the data is decrypted and again encrypted at the WAP gateway.

WTLS Internal Architecture
The WTLS Record Protocol is a layered protocol which accepts raw data from the upper layers to be transmitted and applies the selected compression and encryption algorithms to the data. Moreover, the Record Protocol takes care of the data integrity and authentication. Received data is decrypted, verified and decompressed and then handed to the higher layers.

The Record Protocol is divided into four protocol clients. The protocol stack is shown in Figure 17. The different clients are described in the following sections. The application protocol is not described here, since it is the interface to the upper layers. [6]

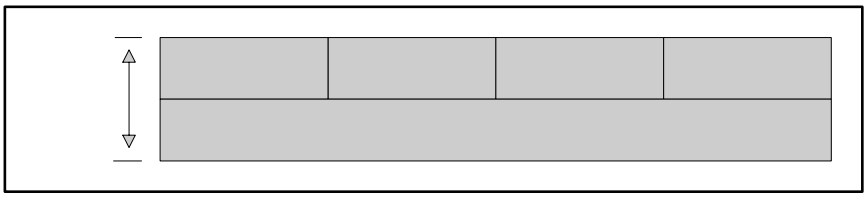

Figure 17 WTLS Internal Architecture

The Handshake Protocol
Every communication between the client and the server starts with a handshake where the parties agree on the session capabilities. For example, all security related parameters are agreed on during the handshake. These parameters include attributes such as protocol versions, cryptographic algorithms, information on the use of authentication and public key techniques to generate a shared secret.

The Alert Protocol
The Record Protocol also provides a content type of alert messages. There are three types of alert messages: warning, critical, and fatal. Alert messages are sent using the current secure state, i.e. compressed and encrypted, or under null cipher spec, i.e. without compression or encryption.

If the alert message, labelled as fatal, is sent, then both parties terminate the secure connection. Other connections using the secure session may continue but the session identifier must be invalidated so that the failed connection is not used to establish new secure connections.

A critical alert message results in termination of the current secure connection. Other connections using the secure session may continue and the secure identifier may also be used for establishing new secure connections.

Error handling in the WTLS is based on the alert messages. When an error is detected the detecting party sends an alert message containing the error that occurred. Further procedures depend on the level of the error that occurred. [6]

The Change Cipher Spec Protocol
The Change Cipher Spec is sent to peer either by the client or the server. When the Change Cipher Spec message arrives, the sender of the message sets the current write state to the pending state and the receiver also sets the current read state to the pending state. The Change Cipher Spec message is sent during the handshake phase after the security parameters have been agreed on.

6.3.2 Authentication

Authentication in the WTLS is carried out with certificates. Authentication can occur between the client and the server or the client only authenticates the server. The latter procedure can happen only if the server allows it to occur. The server can require the client to authenticate itself to the server. However, the WTLS specification defines that authentication is an optional procedure.

6.3.3 Key Exchange

In order to ensure a secure communication channel, encryption keys or initial values to calculate keys have to be exchanged in a secure manner. It is possible that the Server Certificate Message did not contain enough data to allow client to exchange the pre-master secret (pre-master secret is an initial value which is used to calculate the master secret). In this case a Server Key Exchange message is used to provide such data.

The key exchange mechanism of the WTLS also provides an anonymous way to exchange keys. In this procedure, the server sends a Server Key Exchange message which contains the public key of the server. The key exchange algorithm may be RSA [RSA], Diffie-Hellman [DH1], or the elliptic curve Diffie-Hellman [ECDH].

With both RSA and anonymous RSA the client encrypts pre-master secret with the server's public key and sends it back to the server in the Client Key Exchange message. With Diffie-Hellman based algorithms the client and the server calculate the pre-master secret based on one of the private key and the counterpart's public key.

If the client has listed the cryptographic key exchange methods, which it supports, the server may choose whether it is going to use client's suggestions or define another method. If the client has not proposed any method the server has to indicate them.

6.3.4 Confidentiality

Confidentiality in the WTLS is implemented by means of encrypting the communication channel. The encryption methods and all the necessary values for calculating the shared secret are exchanged during the handshake.

The most common bulk encryption algorithms are supported such as RC5 [RC5] with 40, 56 and 128 bit keys, DES [DES] with 40 and 56 bit keys, 3DES [3DES], and IDEA [IDEA] with 40, 56 and 128 bit keys. They are all block cipher algorithms, no stream ciphers except NULLs are supported.

6.3.5 Integrity

Data integrity is ensured using the message authentication codes (MAC). The MAC algorithm is decided at the same time as the encryption algorithm. The client sends a list of supported MAC algorithms where the preferred algorithm is the first in the list.

The WTLS supports common MAC algorithms, such as SHA [SHA] and MD5 [MD5]. There are several different versions of both algorithms, e.g. SHA exists with 0, 40 and 80 bit MAC sizes. The keyed MACs are calculated using the SHA-1. The modified algorithms are based on the SHA-1 but only part of the output is used. Same kinds of versions exist of the MD5 algorithm.

A special MAC algorithm is the SHA_XOR_40 which is a 5-byte checksum. First the input data is divided into the 5-byte blocks. Then all blocks are XOR'ed one after another. It is required that the XOR MAC must be encrypted and is only used for CBC (Cipher Block Chaining) mode block ciphers. The algorithm is intended for devices with limited CPU resources. The MAC is generated over the compressed WTLS data.

6.4 Known Security Holes

The WTLS specification has been adopted from the TLS specification with some modifications and changes. These modifications and changes have led to some security problems, the chosen plaintext data recovery attack, the datagram truncation attack, the message forgery attack and the key-search shortcut for some exportable keys. [13]

Initial vectors called IVs are used by the CBC mode block ciphers to create entropy. Entropy is needed to protect the symmetric key that is used in the CBC mode block cipher. Without IV, the original plain text would be encrypted with a master key. This would open a possibility to use brute force methods to find the shared secret. The usage of IV prevents this happening because the first block in the packet is first XOR'ed with IV. Knowing the content of the original packet does not help in any way, because it is XOR'ed.

Because the WTLS supports an unreliable datagram support where datagrams may be lost, duplicated, or reordered, the CBC mode needs a new IV for encrypting each packet. The IV is computed by XOR'ing the sequence number of the packet and the original IV, which is derived during the key generation. This is also called a linear IV computation. The first plaintext block in the packet is then XOR'ed with the computed IV. The original IV is computed based on values sent during the handshake. All these values are sent without encryption, so they can be eavesdropped. These predictable IVs lead to chosen-plaintext attacks against low-entropy secrets. This security problem affects privacy.

The unauthenticated alert messages, used in the WTLS, let the active attacker replace an encrypted datagram with an unauthenticated plaintext alert message with the same sequence number without being detected. This security problem affects integrity.

For the man-in-the-middle attack, an attacker must actively change one or more handshake messages. If this occurs, the client and the server will compute different values for the handshake message hashes. As a result, the parties will not accept each others' finished messages. Without the master_secret, the attacker cannot repair the finished messages, so the attack will be discovered.

7 End to end security, based on WAP transactions

This chapter describe end-to-end security based on protecting WAP-transactions. There are shown two different solutions and one combination of the both.

By end-to-end security means the security of the communication between the terminal and the server containing the service. When a service is accessed with a WAP terminal, the server containing the service might not necessarily be in the operator network. The server might be accessed through the Internet. GSM or GPRS security mechanisms are not enough for providing end-to-end security for WAP applications. Additional security mechanisms need to be implemented in the mobile terminals and the network. [6, 17]

7.1 Background

WAP network architecture consists conceptually of WAP clients, WAP gateways and content servers.

- A *WAP client* is typically a small handheld device such as a mobile phone.
- A *WAP gateway* is a network element that acts as an intermediate between WAP clients and content servers. The gateways are accessed through a Network Access Point (NAP), where clients dial in.
- A *content server* is a device that contains content or creates it when requested by a WAP client, normally it is an HTTP server. The clients communicate with WAP gateways using WAP protocols. The gateways communicate with content servers normally by using IP-protocols.

Normally an operator maintains a WAP gateway, which its customers use. All content servers accessible through a WAP gateway do not need to belong to the operator. A party that operates a content server and proves services to WAP clients is referred from now on as a third party content provider.

A typical WAP transaction is a request-response transaction, where the client sends request to a content server, which in turn returns a response. First the client makes a WAP protocol request that it sends to a WAP gateway. Then the gateway translates the request into an HTTP request and sends it to a content server. The content server sends the requested content to the WAP gateway and the gateway translates the data into WML binary format and sends it back to the WAP client.

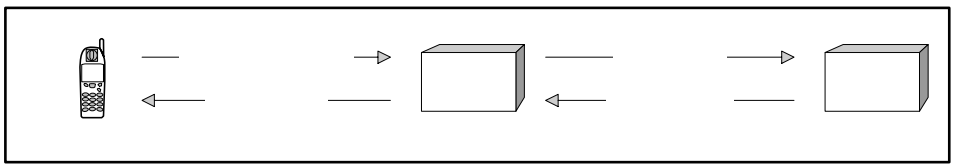

Figure 18 WAP transaction

The security in this kind of WAP transaction is based on the security of underlying IP and GSM/GPRS networks. IP networks have no built-in security and the traffic between the WAP gateway and the content server can easily be eavesdropped and tampered with. In addition, GSM/GPRS has some security vulnerabilities. The security of GSM is based on the encryption used in the air interface, between the MS and BSC. In GPRS data is encrypted between MS and SGSN, no encryption is used be default in the bearer network. This means that WAP transactions without additional security mechanisms are insecure.

However, there are mechanisms to make insecure communication more secure. Transport Layer Security (TLS) protocol is a generic security protocol that can be used to secure any application protocol on top of TCP. It allows HTTP applications to communicate in a way that is designed to prevent eavesdropping, tampering and message forgery. TLS also provides client and server authentication using certificates.

Similarly like HTTP can be secured with TLS, WAP can be secured using Wireless Transport Layer Security protocol (WTLS) specified in the WAP protocol suite. WTLS got functions equivalent to TLS with optimisations for low-bandwidth bearer networks and terminals and which have limited CPU power and memory.

This makes it possible to achieve some security in the WAP transactions when communicating between the WAP client and the content server. If the HTTP communication between the gateway and the content server is encrypted using TLS, and WTLS is used between the client and the gateway. Even with these security mechanisms, the data is still decrypted and encrypted at the WAP gateway. Regardless of using WTLS and TLS, a malicious operator can eavesdrop and tamper with the data, which causes a security threat.

WAP specifications state that if end-to-end security is desired, then the communicating parties must communicate directly using WAP protocols. This implies that content servers, who want to take end-to-end security in use, must support WAP protocols, which they normally don’t do. A common way for content providers is to have a normal HTTP server as a content server.

7.2 Existing Security Solutions

Some level of end-to-end security can be achieved using existing security protocols and dedicated network access points, where WAP users can dial in.

7.2.1 Two Secure Channels Approach

One way to provide end-to-end encryption from the WAP client to the content server is to have two separate secure channels. The first one is a WTLS channel from the WAP client to the WAP gateway and the second one is a TLS channel from the WAP gateway to the content server.

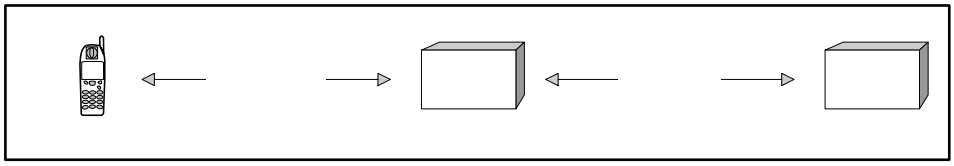

Figure 19 Two Secure Channels

One problem with this method is that the communicating parties, the WAP client and the content server, cannot actually know whether there is end-to-end encryption. The WAP client uses encryption when communicating with the gateway, but it does not know whether encryption is used between the gateway and the content server. In addition, the content server which uses TLS with the gateway, does not know whether the gateway and the WAP client use WTLS.

The traffic flow between the WAP client and the content server is decrypted, and then encrypted again at the WAP gateway of the operator. Allowing a malicious operator to modify and eavesdrop on the data.

So, with this solution, data integrity and confidentiality properties are not achieved by the two secure channels approach. This approach does not either achieve the non-repudiation property.

7.2.2 Using an own Network Access Point

Another solution for the content provider to achieve end-to-end security is to have its own WAP gateway and Network Access Point (NAP), where the users dial in. WTLS may be used between WAP clients and the WAP gateway, and TLS between the WAP gateway and the content server.

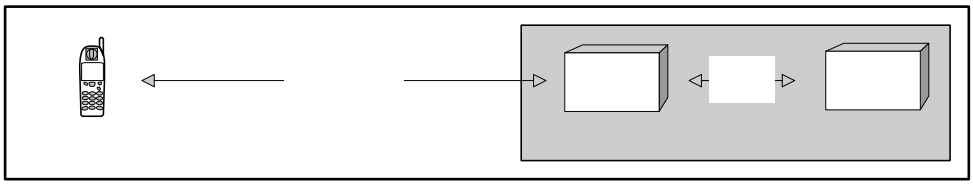

Figure 20 Contents Providers with it's own WAP Gateway

This solution eliminates the problems related to trusting the operator. However, the main disadvantage is that the users have to maintain a list of phone numbers for each content provider that they wish to access. The content providers must in turn bear the extra costs involved in maintaining a WAP gateway.

The approach fulfils the integrity and confidentiality properties but does not achieve the non-repudiation property.

7.3 WAP Transport Layer End-to-End Security

WAP transport layer end-to-end security is a specification by WAP Forum for providing end-to-end security. It allows WAP clients to establish a WTLS connection straight with a WAP gateway of a third party content provider. This can be done even if the WAP client accesses the network through the network access point and WAP gateway of an operator. [6]

The transport layer end-to-end security specification defines new functionality to WAP clients, WAP gateways and content servers. When a client requests some content from a third party content provider, the request is first delivered through the gateway of the operator to the content server of the operator. The content server notices that it does not have the requested content and tells the WAP gateway where the content can be found. The gateway forwards this information to the WAP client, which establishes a WTLS connection with the WAP gateway of the third party content provider. After this, the content can be fetched as usual.

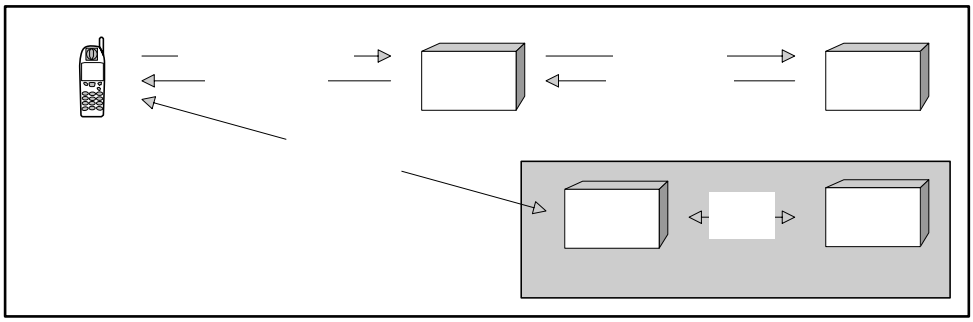

Figure 21 WAP Transport Layer End-to-End Security

The WAP gateway of the operator no longer decrypts and encrypts the traffic destined to a third party content provider. Therefore, a malicious operator can not break the integrity and confidentiality of the data. However, without some security mechanism the integrity and confidentiality are not guaranteed in the communication between the WAP gateway of the third party content provider and the content server. TLS or some lower level mechanism can be used to provide sufficient protection. Thus, it can be stated that integrity and confidentiality properties are fulfilled.

In this approach every content provider must have a WAP gateway of its own, which is costly and awkward. Beforehand agreement is also needed between the operator and every third party content provider, which want to use end-to-end security. This is because the WAP gateway of the operator decides which content providers can be secured using transport layer end-to-end security.

Non-repudiation property is not achieved in this solution. But WAP Forum has specified SignText function as a part of WMLScript. With SignText, a WAP client can digitally sign the transaction data that is sent to the content server. Full non-repudiation is not achieved, because SignText only allows the WAP client to sign a transaction. In that way the WAP client can not deny having done a transaction, but the content server can, because it has not signed anything.

SignText can also be used without the WAP transport layer end-to-end security solution. In this approach, a digital signature in the transaction data provides the non-repudiation and data integrity properties, when confidentiality is not provided. The non-repudiation in this approach has the same problem as mentioned above.

SignText introduces unfortunately new problems. Digital signature keys need to be somehow distributed to all WAP clients that wish to use SignText.

8 Conclusion

This chapter gives first an evaluation of the different technologies discussed in this paper, and finishing with a selected combination.

8.1 Conclusion in Bluetooth

In every Bluetooth device, there are four entities to maintain the security at the link level; the Bluetooth device address (BD_ADDR), a private authentication key, a private encryption key, and a random number referred as RAND.

There are three security modes for a device:
- *Security Mode 1* (non-secure): A device will not initiate any security procedure.
- *Security Mode 2* (service level enforced security): A device does not initiate security procedures before channel establishment at L2CAP level.
- *Security Mode 3* (link level enforced security): A device initiates security procedures before the link set-up at the Link Manager Protocol (LMP) level is completed.

In addition, two trust levels are distinguished for devices:
- *Trusted Device:* Device with fixed relationship, which is trusted and has unrestricted access to all services.
- *Untrusted Device:* Device that is not considered as trusted and has restricted access to services.
- *Unknown Device:* No security information is available for this device, which makes it to an untrusted device.

For services the requirement for authentication, authorisation and encryption are set independently. The access requirements allow three security levels to be defined:
- *Services that require authentication and authorisation.* Automatic access is only granted to trusted devices. Other devices need manual authorisation.
- *Services that require authentication only.* Authorisation is not necessary.
- *Services open to all devices*; authentication is not required, no access approval required before service access is granted.

As described, there are many different security combinations and level of security to choose among. The highest level on a device uses both authentication and encryption. This is done with the HCI commands HCI_Write_Encryption_Enable and HCI_Write_Authentication_Enable. These values are persistent for all new connections and are not possible to change remotely. However, applications that run on a Bluetooth device may set the level of security that is required.

Even with all these defined security levels of devices and services, the use of application level key exchange and encryption methods to secure the communication on top of the existing Bluetooth security systems are recommended. This because the use of a normal

PIN code in the initialisation process will be too weak, small and inconvenient to remember and use every time when connection to a device.

Without regard to which level of security that is used, none of them prevents the optional use of higher layer security mechanisms. This may be VPN, IPSEC, TLS/WTLS, application level security, etc.

8.2 Conclusion of GSM/GPRS security

By default, neither GSM nor GPRS provide any encryption beyond the air interface; this means that the operator does not provide end-to-end encryption in GSM/GPRS.

The main security functions related to the GPRS device (MS) are authentication and encryption. Authentication is performed in the same way as in GSM with a challenge-response protocol. Encryption in GPRS is a bit different from encryption in GSM, and is done between the MS and the SGSN, instead of between the MS and the base station (BSC) as in GSM.

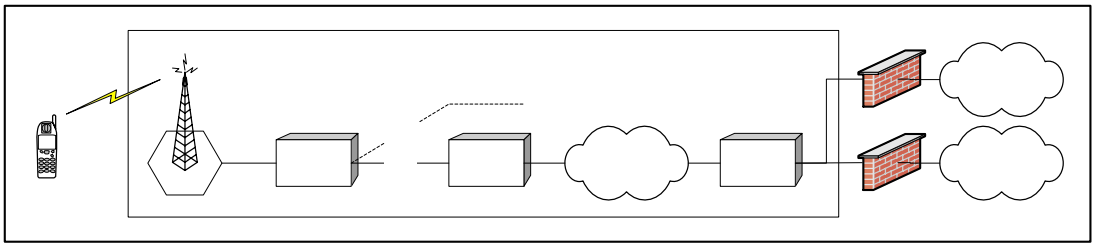

Figure 22 The Operator's GPRS Network

GPRS backbone security is ensured by restricting access to the network, and by protecting the network boundaries by placing firewalls between the GPRS network and networks belonging to other operators as well as other external networks.

Inter-PLMN links are secure providing integrity and confidentiality. For example, secure links can be achieved by point to point links, private Inter-PLMN backbones or encrypted tunnels over the public Internet.

Some weaknesses in 2G Security is that it is only the network that can initiate the encryption, and it is an option, not mandatory. Nor get the user any indication if encryption is provided or not. Moreover, it is only the networks that authenticate the subscriber, the subscriber cannot authenticate the network.

In 3G Security both the operator and the user have to authenticate themselves to each other, and there will be an indicator that tells if encryption is in use or not.

The conclusion of the 2G Security in GSM/GPRS is that the users have to trust the operator. Moreover, to get secure data communication there has to be supplementary security mechanisms in addition to what the GSM/GPRS provides.

8.3 Conclusion in IPSec

IPSec is a pair of protocols, ESP (for Encapsulating Security Payload) and AH (for Authentication Header), which provide security services for IP datagrams.

IPSec introduces the concept of a Security Association (SA). A SA is a logical connection between two devices transferring data. It provides data protection for unidirectional traffic by using the defined IPSec protocols. It is uniquely identified by a triple consisting of a Security Parameter Index (SPI), an IP Destination Address, and a security protocol (AH or ESP) identifier. An IPSec tunnel consists normally of two unidirectional SAs, which together provide a protected, full-duplex data channel.

ESP provides authentication, integrity, replay protection, and confidentiality of the data, in other words it secures everything in the packet that follows the IP header. While AH provides authentication, integrity, and replay protection (but not confidentiality). The main difference between AH and ESP, is that AH also secures parts of the IP header of the packet. Nevertheless, ESP encrypts the payload data with an encryption algorithm using a secret encryption key. Only the ones knowing this key can decrypt the data, thus providing confidentiality. Both the algorithm and the encryption key are parameters of the SA.

IPSec operates in two modes, either tunnel or transport mode. In transport mode, the ordinary IP header is used to deliver the packets to their endpoint. In tunnel mode, the ordinary IP header only tells the address of a security gateway, knowing how to verify/decrypt the payload and forward the packet to a destination given by the IP header contained in the protected payload.

To secure two networks or only two devices with IPSec, the conclusion would be to use ESP even though this gives the most overhead, it still gives the best security. Use tunnel mode to secure two networks, in that way the whole packet is encrypted. And transport mode to secure two devices, since the destination address already is known when communicating between two devices.

8.4 Conclusions of the WTLS protocol in WAP

The Wireless Transport Layer Security protocol is the first attempt to provide a secure end-to-end connection for the Wireless Application Protocol.

Allowing anonymous connection to be established can be very risky, and may lead to man-in-the-middle attacks. The conclusion here is that there should always be mutual authentication, other ways the connection will be shut down. To attain this, the client should define that during the handshake it will not support key exchange suites without authentication. The client wants to be sure that the content server he is communicating with is the one it claims to be. In addition, the server wants to be sure it is the right client that got the services it has paid for and not anyone else.

WAP transactions in GSM/GPRS networks are insecure unless additional security mechanism is used. Some of these problems may be solved with the use of WTLS in the Wireless Application Protocol.

Two Secure Channels Approach
One problem with this method is that the communicating parties, the WAP client and the content server, cannot actually know whether there is end-to-end encryption. The WAP client uses encryption when communicating with the gateway, but it does not know whether encryption is used between the gateway and the content server. In addition, the content server which uses TLS with the gateway, does not know whether the gateway and the WAP client use WTLS.

With this solution, the traffic flow between the WAP client and the content server is decrypted and then encrypted again at the WAP gateway of the operator. Allowing a malicious operator to modify and eavesdrop on the data.

Using an own Network Access Point
This solution eliminates the problems related to trusting the operator. However, the main disadvantage is that the users have to maintain a list of phone numbers for each content provider that they wish to access. The content providers must in turn bear the extra costs involved in maintaining a WAP gateway.

WAP Transport Layer End-to-End Security
When a client requests some content from a third party content provider, the request is first delivered through the gateway of the operator to the content server of the operator. The content server notices that it does not have the requested content and tells the WAP gateway where the content can be found. The gateway forwards this information to the WAP client, which establishes a WTLS connection with the WAP gateway of the third party content provider.

With this solution, every content provider must have a WAP gateway of its own, which is costly and awkward. Beforehand agreement is also needed between the operator and every third party content provider, which want to use end-to-end security. This is because the WAP gateway of the operator decides which content providers can be secured using transport layer end-to-end security.

Therefore, when communicating using WAP transactions there should always be mutual authentication, and to be sure of the security, there is need of an own WAP gateway. As mentioned earlier, an own WAP gateway is not what a content provider wants. The normal way to do WAP transactions, is with the Two Secure Channels Approach. The conclusion in normal day life must therefore be – trust the operator.

8.5 Combining the technologies

A scenario is a nice way to combining the technologies described in this paper. If a GSM/GPRS telephone got a data connection with a Bluetooth-network thru the Public Land Mobile Network (PLMN) and the Internet, there are several technologies involved.

From the mobile side there are the generally GSM/GPRS security, which provide authentication and encryption on the air link level. The operator also provides secure connections inside its network in form of integrity and confidentiality, but not encryption.

From the other side of the connection there is the Bluetooth security. Bluetooth security provides a vide range of security levels on the services as well as the devices. With the proper use of levels and PIN codes, Bluetooth may provide a reasonable level of security. Although the recommendation of application link level key exchange and encryption methods still stands.

Between the GSM/GPRS operators network and the gateway to the Bluetooth network there also have to be some kind of security. This may be IP Security or WTLS provided by WAP, where they both provide some kind of end-to-end security.

When using a WAP enabled telephone or a Bluetooth enabled device to surf on the Internet or download some files, there are normal security to take care of the data. However, when there is more sensitive information that should be transmitted over the public Internet, as financial transactions, additional security systems are recommended.

IP Security may be used between the operators and the Bluetooth gateways to make only the path between secure. Alternatively between the mobile and the Bluetooth device to provide end-to-end security. Between the gateways, IPSec can be used as a secure tunnel in Tunnel Mode, or between the devises in Transport Mode.

To use the Wireless Transport Layer Security provided by WAP, the communication have to run on top of the Wireless Application Protocol. This security protocol provides a secure communication between the WAP device and the WAP gateway. At the gateway,

the data is decrypted and encrypted again, before the data continues to the content provider. To solve this problem, the content provider has to get his own WAP gateway, which need maintenance and cost money. The normal way to communicate is thru the WAP gateway of the operator, where the content server is a HTTP server. The conclusion of the WTLS end-to-end security is that the users have to trust the operators, and normally is that not a problem.

With IPSec, it is possible to use the mode called Tunnel Mode to secure the communication between two trusted networks. This may be very useful for trusted Bluetooth networks that communicate together via the Internet. Alternatively, between devices with use of the mode called Transport Mode.

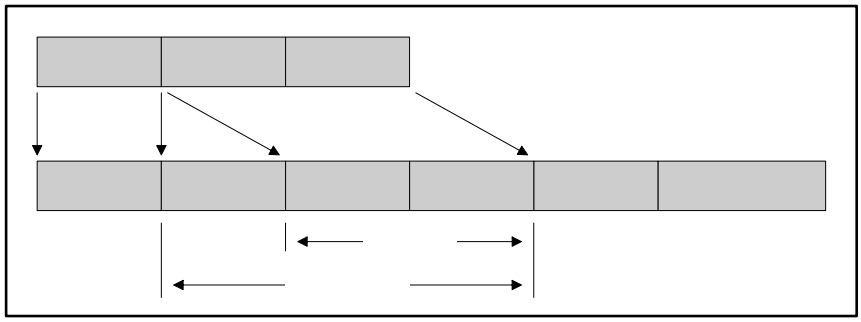

Figure 23 IPSec with ESP header in Transport Mode

If there is no trust in the operator, there may be a solution with use of the IPSec in Transport Mode applied by the ESP header. In this mode, everything after the ESP header is protected by the services ESP are providing; authentication/integrity, replay protection and confidentiality. I would recommend this solution for secure data communication.

9 Abbreviations

A3	Authentication algorithm
A8	Ciphering key generating algorithm
ACL	Access Control List
ACL	Asynchronous Connectionless
ACO	Authenticated Ciphering Offset
AH	Authentication Header
AuC	Authentication Centre
BD_ADDR	Bluetooth Device Address
BG	Border Gateway
BSC	Base Station Controller
BTS	Base Transceiver Station
CA	Certification Authority
CBC	Cipher Block Chaining
CDC	Certification Distribution Centre
CIA	Confidentiality, Integrity and Authentication
CN	Core Network
COF	Ciphering Offset Number
CRC	Cyclic Redundancy Check
DES	Data Encryption Standard
DH1	Diffie-Hellman
DHCP	Dynamic Host Configuration Protocol
ECDH	Elliptic Curve Diffie-Hellman
ESP	Encapsulating Security Payload
EXCH	Exchange (telecommunication switch)
FEC	Forward Error Correction
FHSS	Frequency-Hopping-Spread-Spectrum
FSM	Finite State Machine
FTP	File Transfer Protocol
FW	Fire Wall
GAP	Generic Access Profile
GEA	GPRS Encryption Algorithm
GFSK	Gaussian Frequency Shift Keying
GGSN	Gateway GPRS Support Node
GMSC	Gateway MSC
GPRS	General Packet Radio Service
GSIM	GSM SIM
GSM	Global System for Mobile Communication
GTP	GPRS Tunnelling Protocol
HCI	Host Controller Interface
HLR	Home Location Register
HTTP	Hyper Text Markup Language

ICMP	Internet Control Message Protocol
IDEA	International Data Encryption Algorithm
IEEE	Institution of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IKE	Internet Key Exchange
IMEI	International Mobile Equipment Identity
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
IPSEC	Internet Protocol Security
IWF/IWU	Inter-working Function / Inter-working Unit
Kc	Encryption key
KDC	Key Distribution Centre
Ki	Subscriber Identification Key
L2CAP	Logical Link Control and Adaptation Protocol
LFSR	Linear Feedback Shift Register
LFSR	Linear Feedback Shift Register
LM	Link Manager
LMP	Link Manager Protocol
MAC	Message Authentication Codes
MD5	Message Digest Algorithm
ME	Mobile Equipment
MS	Mobile Subscriber
MSC	Mobile Services Switching Centre
MTU	Maximum Transmission Unit
OBEX	Object Exchange Protocol
OSI RM	Open Systems Interconnection Reference Model
OSI	Open Systems Interconnect
PAN	Personal Area Network
PCU	Packet Control Unit
PDA	Personal Digital Assistant
PIN	Personal Identification Number
PKI	Public Key Infrastructure
PLMN	Public Land Mobile Network
PPP	Point to Point Protocol
QoS	Quality of Service
RADIUS	Remote Authentication Dial-In User Service
RAND	Random Number
RC5	Encryption Algorithm
RF	Radio Frequency
RFCOMM	Radio Frequency Communication
RSA	Rivest, Shamir and Adleman
SA	Security Association
SAFER+	Secure And Fast Encryption Routine
SAR	Segmentation and Reassembly
SAT	Secure Audit Trail

SCO	Synchronous Connection-Oriented
SGSN	Serving GPRS Support Node
SIG	Special Interest Group
SIM	Subscriber Identity Module
SKE	Server Key Exchange
SRES	Signed Response
SPI	Security Parameter Index
SSL	Secure Socket Layer
TCP	Transmission Control Protocol
TE	Terminal Equipment
TLS	Transport Layer Security
TTP	Trusted Third Party
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunication System
VLR	Visitor Location Register
VPN	Virtual Private Network
WAE	WAP Application Environment
WAP	Wireless Application Protocol
WML	Wireless Markup Language
WSP	Wireless Session Protocol
WTA	Wireless Telephony Applications
WTLS	Wireless Transport Layer Security
WTP	Wireless Transport Protocol
XML	Extensible Markup Language
XOR	Exclusive-OR

10 References

[1]	Riku Mettala, White Paper, Bluetooth Protocol Architecture, 25.Aug. 1999, [Referred 2002-02-21] < http://www.bluetooth.org/member/docs/WhitePaper_Protocol_Architecture.pdf >
[2]	L. Zhou & Z. J. Haas, Securing Ad Hoc Networks, IEEE Network Magazine, vol. 13, no.6, November/December 1999, [Referred 2002-02-21] < http://wnl.ece.cornell.edu/ >
[3]	Specification of the Bluetooth System, Core, v.1.1, [Referred 2002-02-18] < http://www.bluetooth.com/pdf/Bluetooth_11_Specifications_Book.pdf >
[4]	Specification of the Bluetooth System, Profiles, v.1.1, [Referred 2002-02-18] < http://www.bluetooth.com/pdf/Bluetooth_11_Profiles_Book.pdf >
[5]	Gollmann D., Computer Security, John Wiley & Sons Ltd., 1999, 336p. ISBN 0-471-97844-2
[6]	WAP Forum, WAP Specifications, [Referred 2002-03-19] < http://www.wapforum.org >
[7]	J. Bray & C. F. Sturman, Bluetooth 1.1: Connect Without Cables, 2nd edition, 2002, 537p. ISBN 0-13-066106-6
[8]	S. Jormalainen & J. Laine, Helsinki University of Technology, Security in the WTLS, 10. Jan. 2000, [Referred 2002-03-19] < http://www.hut.fi/~jtlaine2/wtls/ >
[9]	ETSI EN 301 344, Digital cellular telecommunications system, General Packet Radio Service (GPRS), Service description, V7.4.1, 2000
[10]	Smartpipes Inc, IPSec-Based VPNs, [Referred 2002-04-10] http://www.bitpipe.com/data/detail?id=991168448_785&type=RES&x=1633498327
[11]	ETSI TS 121 133, Universal Mobile Telecommunications System, 3G Security, Security threats and requirements, V4.0.1, 2001
[12]	N. Borisov, I. Goldberg and D. Wagner, Security of the WEP algorithm < http://www.isaac.cs.berkeley.edu/isaac/wep-faq.html >
[13]	M. J. Saarinen, Attacks against the WAP WTLS Protocol, [Referred 2002-04-26] < http://www.jyu.fi/~mjos/wtls.pdf >
[14]	Bluetooth Security Architecture, White Paper, Version 1.0, July 1999
[15]	Nokia, WAP on Web, [Referred 2002-04-29] < http://www.nokia.com/wap/ >
[16]	Cisco Systems, GPRS White Paper, [Referred 2002-05-08] < http://www.cisco.com/warp/public/cc/so/neso/gprs/gprs_wp.htm >
[17]	J. Mynttinen, Helsinki University of Technology, End-to-end security of mobile data in GSM, 27. Nov. 2000, [Referred 2002-05-08]

[18]	ETSI TS 133 120, Universal Mobile Telecommunications System, 3G Security, Security Principles and Objectives, V4.0.0, 2001
[19]	G. M. Køyen, Telenor R&D Agder, Security aspects, 1999
[20]	3GPP TS 03.20, Digital cellular telecommunications system (Phase 2+), Security related network functions, V8.1.0, 2000
[21]	S. M. Bellovin, AT&T Research, Problem Areas for the IP Security Protocols, [Referred 2002-05-20] < http://www.research.att.com/~smb/papers/badesp.pdf >
[22]	N. Ferguson and B. Schneier, Counterepane Internet Security Inc., A Cryptographic Evaluation of IPSec, [Referred 2002-05-20] < http://www.counterpane.com/ipsec.pdf >