"SECURE M-COMMERCE APPLICATIONS IN BANKING INDUSTRY"

**A Dissertation Submitted to Jawaharlal Nehru University
in partial fulfillment of the requirements for
the award of the degree of**

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE & TECHNOLOGY

By

RAJEEV KUMAR SINGH

Under Supervision of
Prof. P.C. SAXENA

School of Computer and Systems Sciences
JAWAHARLAL NEHRU UNIVERSITY

New Delhi - 110 067, INDIA
July. 2005.

School of Computer & Systems Sciences
JAWAHARLAL NEHRU UNIVERSITY
New Delhi - 110067

Certificate

This is to certify that the dissertation entitled "**SECURE M-COMMERCE APPLICATIONS IN BANKING INDUSTRY**" being submitted by Mr. Rajeev Kumar Singh, to School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi for the award of **Master of Technology** in *Computer Science and Technology*, is a record of bonafide work carried out by him under my supervision.

This work has not been submitted in part or full to any university or institution for the award of any degree.

Rajeev Kumar Singh

(Student)

04/08/05

Dean
School of Computer
& Systems Sciences

Prof. P. C. Saxena 13.7.05

(Supervisor)

Acknowledgement

With a deep sense of gratitude, I wish to express my sincere thanks to my supervisor, Prof. P.C. Saxena, for his immense help in planning and executing this master's dissertation work on time. Thanks to him for his constant encouragement, guidance and affection throughout my work

I am grateful to Prof. Saxena for providing me enough infrastructures through Data Communication and Distributed Computing Group (DCDCG) laboratory to carry out my work. I am thankful to all the members of the group for their co-operation, encouragement and understanding

I would like to thank all my faculty members for their help and useful suggestions during my stay in JNU. I would also like to thank all my classmates particularly Avinav Dongre and my juniors for their love and affection.

I sincerely thank all, for the help and encouragement

RAJEEV KUMAR SINGH

Abstract

GSM has changed the face of communication and information exchange, much as the Internet did. With the advances made in the mobile technology arena, new opportunities are created. Mobile Commerce (m-Commerce) is one such opportunity. Each new advance in technology brings with it, associated risks. This dissertation focuses on the risks involved with m-Commerce for the banking industry.

This dissertation provides a detailed overview of basic services that any m-Commerce application should provide to the banking industry.

The security of GSM networks has come under attack in the past. This is largely due to the fact that the GSM consortium opted to develop their security technologies in secret, rather than in the public domain. This dissertation aims to evaluate the security offered by GSM and assess potential attacks in order to further understand risks associated with m-Commerce applications over GSM.

The arrival of the SIM Application Toolkit and the Wireless Application Protocol promised to again change the face of commerce. A detailed analysis of these enabling technologies is presented in the dissertation.

Some changes to either the application architectures or the processing of the data have been suggested in order to enhance the security offered by these services. These principles provide the foundation for securing any financial transaction over untrusted networks

CONTENTS

Chapter 1

INTRODUCTION

1.1Background

It's a mobile service provider's dream. In the near future, a man will walk down the street and his mobile phone will vibrate. "ABC Corp. releases third quarter profit warning", says the readout. "Recommendation: Sell ABC stock. To sell, press '1'.To keep press '2'.The man pauses to consider, takes a deep breadth and presses '1'. "Sale executed ", read his handset. Heady from the excitement, he needs a soda to calm his nerves. He goes to a machine and, instead of fishing through his pockets for change, points the handset at the soda machine, keys in a code and pushes a button on the machine .A soda pops out. This dream will be real in a near future due to advances in mobile commerce.

While mobile commerce is not delivering the promises that many pundits had proclaimed just a few years ago in terms of providing unprecedented commercial functionality to the masses, it is still projected to be one of the main driving forces for next generation computing and a major revenue-generating platform for many corporations.

Research firm IDC confirmed US$500 million in m-commerce revenues for 2002 and projects the amount to be US$27 billion by 2005.Forrester Research also predicts that by 2007, up to 2.3 million wired phone subscribers in the US would make the switch to wireless access, making an average of 2.2 wireless phone per household

Keen and mackintosh note that the key value proposition of mobility is the creation of choice, or new freedoms, for customers. In a similar way, words commonly used to describe that main value –added feature of m-commerce include flexibility, convenience, and ubiquity. The distinctive feature of mobile commerce is the significance of the user's location, his situation, and his mission.

The essential features of m-commerce are:

1. **Always on**: Because of its inherent design, a mobile phone can be 'always on' and is always portable. This permits its users to engage in activities such as meeting with people or traveling while conducting transactions through their Internet- enabled mobile devices.

2. **Location-centric**: Not only does a mobile phone go everywhere, GPS may also be constructed to recognize where the phone is and to personalize the available services accordingly. Knowing the location of the Internet user creates a significant advantage for m-commerce over wired e-commerce. Utilizing this technology, m-commerce providers will be able to better receive and send information relative to a specific location.

3. **Convenience**: People will no longer be constrained by time or place in accessing e-commerce activities. Rather, m-commerce could be accessed in a manner that may eliminate some of the labor of life's activities. For example, consumers waiting in line or stuck in traffic will be able to pursue favorite Internet based activities or handle daily transactions through m-commerce applications. Consumers may recognize a special comfort that could translate into an improved quality of life. By making services more convenient the customer may actually become more loyal. Consequently, communication facilities within m-commerce are key applications for the delivery of convenience.

4. **Customization**: Mobile phones have a much higher penetration than PCs, so m-commerce producers can be more creative and customizable in designing segmented, lifestyle tools. For instance, using demographic information collected by wireless service providers and information on the current location of mobile users, more targeted advertising can be done. The advertising messages can be customized based on information provided by consulting the user at an earlier stage or by the history of users` purchasing habits.

5. **Identifiability**: A mobile phone has a built-in ID to support secure transactions whereas a PC is virtually anonymous.

1.2 M-commerce operation modes

M-commerce operation modes can be generalized in two categories: (1) content delivery (notification and reporting) mode and (2) transaction (purchasing and data entry) mode. They are about having access to information and being able to carry out a particular transaction unconstrained by time and location

In the content delivery mode, as shown in Table 1, the mobile web is used to notify and report important content messages such as sports news, personalized financial news, premium games, and mobile greeting cards. All content providers must ensure that their services are optimized for the mobile channel and live up to the highest levels of quality and usability. The centerpieces of the content delivery are personalized information. The personalization involves presenting choices relevant to the time and place of interaction and is based on the user's previous transactions or preferences. For example, a customer who makes a restaurant reservation online could be prompted for a mobile direction map and a choice of post-dinner events, all based on previous behavior. In the transaction mode, companies use the wireless Internet to run business transactions. M-commerce consumers can browse through the catalog and order products online. Although there are still some hidden obstacles (e.g., transaction security, speed, and ease of use), it seems that most companies are likely to benefit directly from transactions on the wireless Internet, especially for small and medium-sized enterprises. In fact, there have been many successful cases, including the On-Pay m-commerce system, which is capable of executing transactions from external online merchants including vending machines, tickets, trains, gasoline, and taxi fares. Time sensitive and simple (yes-or-no) transactions are the key success factors to this operation mode. M-commerce adds mobility and convenience to the Internet and creates a whole new set of opportunities.

1.2.1Table-1 M-commerce operation modes

	Content delivery mode	Transaction mode
Definition / characteristics	This operation notifies and reports important content messages to consumers such as sports news and personalized financial news.	This operation runs business transaction over the wireless Internet. Consumers can browse through the catalog and order products on mobile
Promotion measures / Ways	Sending instant coupon to near-by customers; Notifying customers when they have been outbid in an auction that is about to close; Delivering sports news, financial news, and personalized information; Offering ring-tones and other downloads to m-phone customers; and Using the Web as a cost-effective way to augment its core products with related information and service functions.	Using micro-payment technology in transactions involving vending machines, tickets, trains, or taxi fares; Automating ubiquitous customer billing transaction services; Running real-time ubiquitous online auction transactions; Charging a fee for mobile games, entertainment, and fun; and Providing convenience to implement a transaction at any time or place.

1.3 Problem Statement

Technology comes together with some problems. Security is an essential part of most mobile commerce application. However, security in mobile networks differs fundamentally from the security in fixed networks. This is partly due to the fact that security in mobile networks is to large extent controlled by the operator. As a result, the security between the user and the content or service provider is typically split into two domains, the operator acting as a translator between the mobile and fixed network. This implies that both user and content provider have to trust the operator to handle data and transaction in secure way.

In their article titled "Enhancing Security of GSM" [3] states that although voice data is protected on the radio link between the mobile handset and the GSM Base Station, there are no protection offered during the transmission of data through the fixed network of the mobile communication provider. This fact could lead to eavesdropping of the voice data as `well as tampering with Short Message Service (SMS) messages sent between communication entities.

SMS uses store and forward technology therefore, in [3] author argues that SMS messages are highly accessible for attacks from the network provider side even though they are encrypted over the air link. He also states that injection of malicious and false SMS messages is possible through poorly protected SMS gateways. This indicates that Authentication, Authorization, Confidentiality and Integrity of the SMS communications are at risk.

These factors imply that one cannot blindly trust the security offered by GSM and its accompanying application. An alternative solution is required and this study investigates the possibilities and proposes a secure solution.

1.4 Objectives

The objectives of this dissertation are

1. Understanding the basic principles of information security and how to obtain them
2. Analyzing the security offered by the GSM network of both over-the –air and the fixed network, with special emphasis on message integrity.
3. Analyzing the security offered by wireless Internet gateway (WIG) and wireless application protocol (WAP) application
4. Applying the basic principles of information security to WIG and WAP architecture in order to find and propose enhancement to the security of theses solutions

1.5 Outline

The dissertation follows a logical order keeping in view the interest of the reader. In Chapter 1 the author initiates the discussion of M-commerce with Introduction. In Chapter 2 the basics of security is discussed. The discussion revolves around the basic requirements that information security should achieve. Chapter 3 discusses about GSM architecture and the security provided by GSM. Then it discusses about flaws in GSM security. Chapter 4 discusses looks at the WIG application and its enabling technologies in great detail Chapter 5 an alternative solution is proposed Chapter 6 proposed solution is analyzed Chapter 7 all the strings are tied together in the conclusion

Chapter 2

MESSAGE INTEGRITY

2.1 Introduction

With the introduction of the computer and the advent of computer networks, the need for protecting information became very evident. Company and customer data traversed the open networks and this information fell into the wrong hands or got altered without the knowledge of the author of the message. Automated tools were required for protecting the sensitive data flowing over these networks. Cryptography came as a clear answer to all the concerns.

In this chapter we will look at transaction/ message security over open networks regardless of the transmission media used. Transaction security is critical in any commerce application but even more so in the e-Commerce and m-Commerce domains, as the electronic messages are transmitted over open, unsecured networks.

2.2 Attacks On Electronic Messages

An attacker might want to gain access to an electronic message for numerous reasons. Gaining unauthorized access to information in order to violate someone's privacy, impersonating another user in order to shift the responsibility or originate a fraudulent activity are some of the reasons an attacker might want to access the information.

There are four general categories of attacks on a transmitted message apart from normal transaction flow: [6]

- Interruption.
- Interception.
- Modification.
- Fabrication.

Each one of these will now be discussed in detail.

2.2.1Normal message flow

In general there is a flow of information from a source to a destination. In a normal message flow, the information passes unhindered from the source to the destination as shown in Figure 2.1.

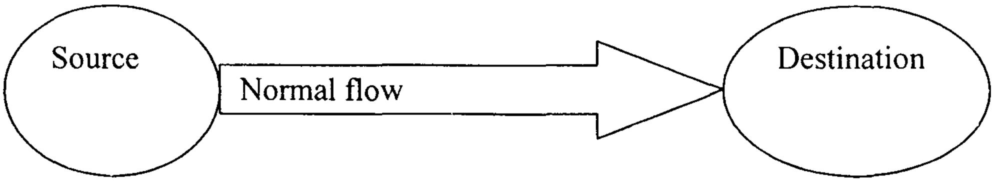


Figure 2.1: Normal message flow

2.2.2 Interruption

Interruption is the action of preventing a message from reaching its intended recipient. It can also occur when an asset of the system is destroyed or becomes unavailable or unusable. This is an attack on availability. Some examples of these kinds of attack include: [6]

- Destruction of a piece of hardware.
- The intentional cutting of a communication line.
- Disabling of the file management system.
- Denial of service attack.

Figure 2.2 illustrates this graphically.

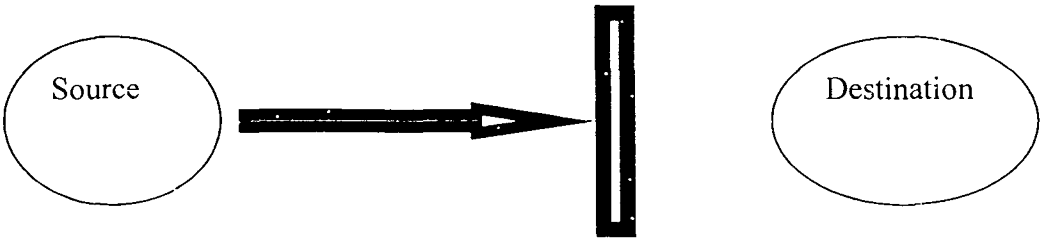


Figure 2.2: Interruption of a message.

2.2.3 Interception

Interception is where an unauthorized party gains access to information. This is an attack on confidentiality [6]. The unauthorized party might be a person, program or a computing system. A loss due to this kind of attack might be noticed quickly, but a silent interceptor might leave no traces by which the interception can be detected. Examples of these kinds of attacks include: [6]

- Wiretapping to capture data in a network.
- Illicit copying of files or programs.

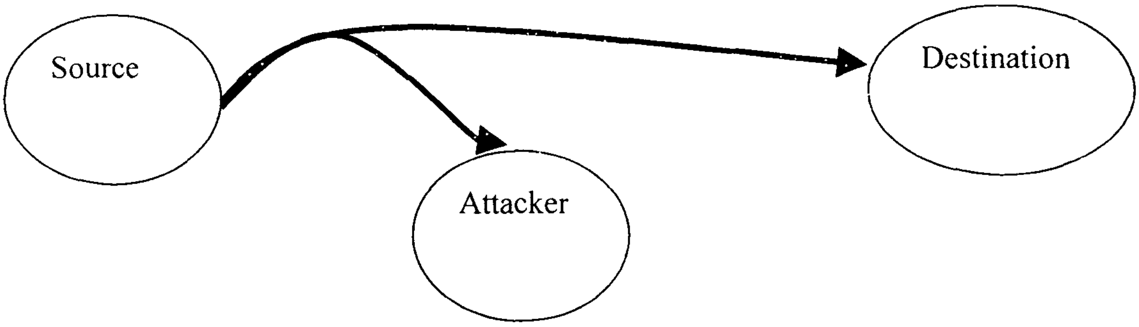


Figure 2.3: Interception of a message. [6]

2.2.4 Modification

Modification is where an unauthorized party not only gains access to an asset, but tampers with it. This is an attack on the integrity of the message. Examples include: [6]

- Changing of values in a database for personal gain.
- Altering a program so that it performs an additional computation.
- Modifying the content of a message transmitted on a network.

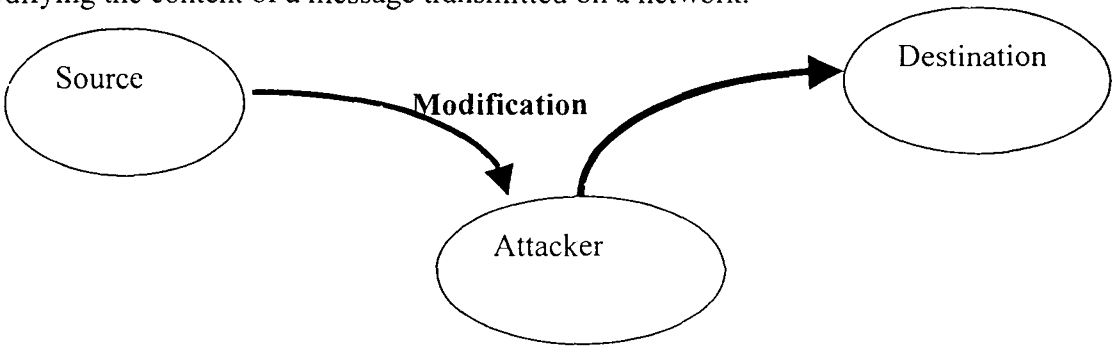


Figure 2.4: Modification of a message. [6]

2.2.5 Fabrication

Fabrication occurs when an unauthorized party inserts counterfeit objects into the computing system. This is an attack on the authenticity of the message [6]. These insertions can sometimes be detected as forgeries, but if skillfully done, they are virtually indistinguishable from the real thing. Examples include: [6]

- Insertion of spurious information into the network communication system.
- Adding additional records to an existing file or database.

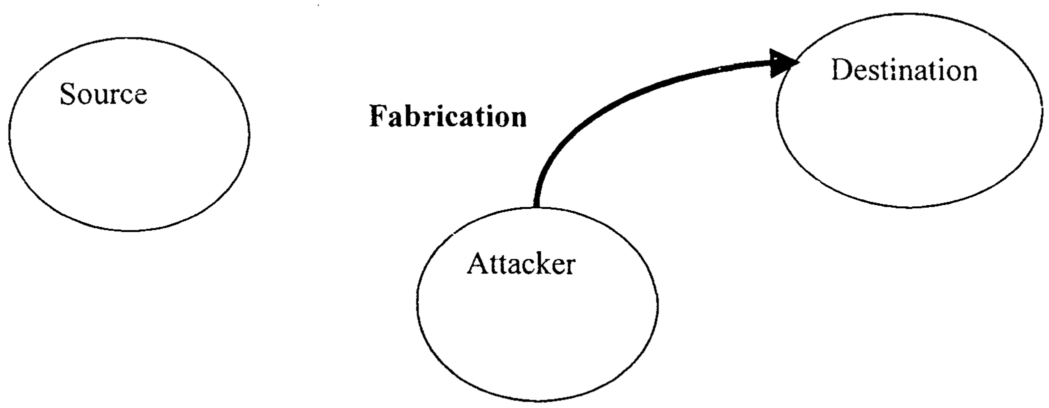


Figure 2.5: Fabrication of a message. [6]

2.3 Objectives Of Security In Message Transmission

In knowing and understanding the attacks on messages we can look at the goals of securing transmissions. Regardless of who is involved, all parties to a transmission must have confidence that certain objectives associated to message transmission have been met. One such means of ensuring this is by means of cryptography.

Cryptography is the study of mathematical techniques related to aspects of information security like confidentiality, message integrity, entity authentication etc.. Cryptography provides the means to ensure that the objectives of communicating parties are met. These objectives are briefly discussed.

Confidentiality: keeping information secret from all but those who are authorized to see it.

Message integrity: Ensuring the transmitted message has not been altered by unauthorized or unknown means.

Authentication: Corroboration of the identity of an entity, like a person, a computer terminal etc. and corroboration of the origin of the message.

Non-repudiation: Preventing the denial of some previous commitments or actions by the communicating parties.

Availability: Availability provides functionality to ensure that resources or information are accessible and usable upon demand by authorized users.

Authorisation: Authorization provides functionality to determine whether users or applications are permitted to use computer resources.

2.4 Ensuring Message Integrity During Transmissions

In any financial transaction, over any untrusted network, it is important to ensure that the integrity of the message is not compromised. To this end we will look at ways to ensure the integrity of messages during transmission over said unsecured networks. It is important to note here that the transmission media is of no concern to the applications listed below.

Figure 2.6 below illustrates the procedure for verifying the integrity of transmitted messages that were protected by use of some cryptographic technique, the result of which is denoted by the term Authentication Code (AC).

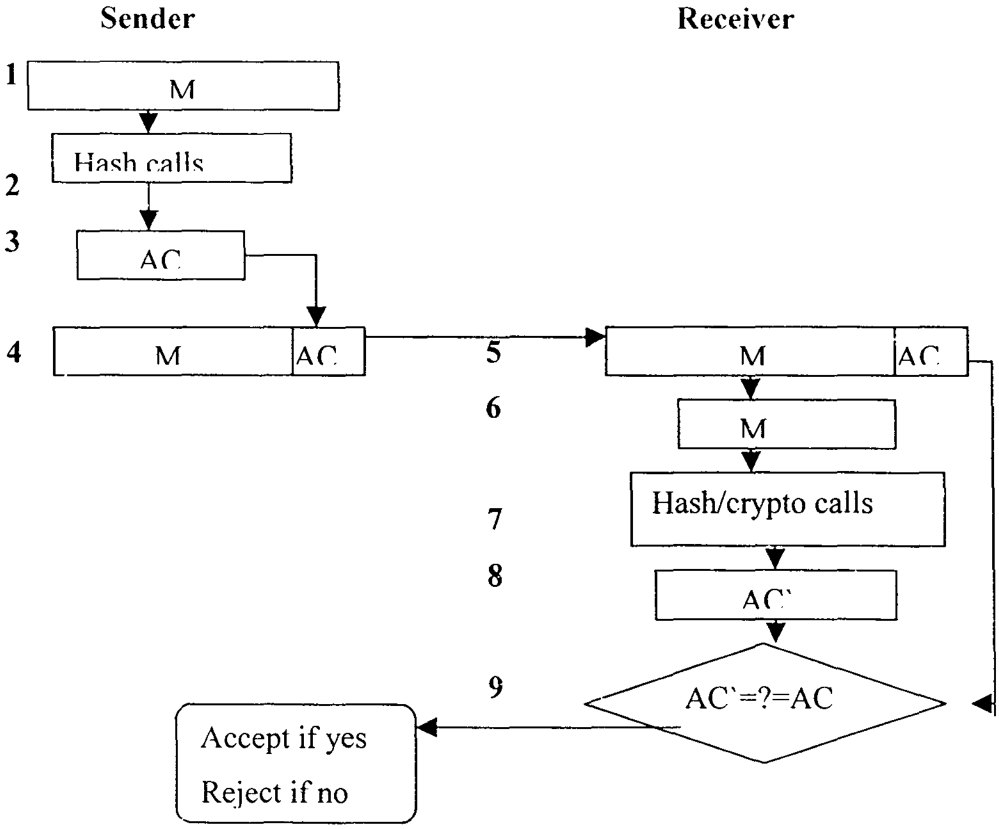


Figure 2.6: Message Integrity Verification Procedure

2.4.1Hash functions

A hash function can be defined as a function that compresses an input string of arbitrary length to an output string of fixed length. There are numerous implementations of such hash functions like the MD4, MD5, HAVAL and SHA-1 algorithms. The current international standard for hash functions is the SHA-1 algorithm and as such when we refer to a hash function in the text, we are referring to the SHA-1 algorithm.

Figure 2.7 graphically illustrates how the SHA-1 algorithm gets applied to a message of arbitrary length.

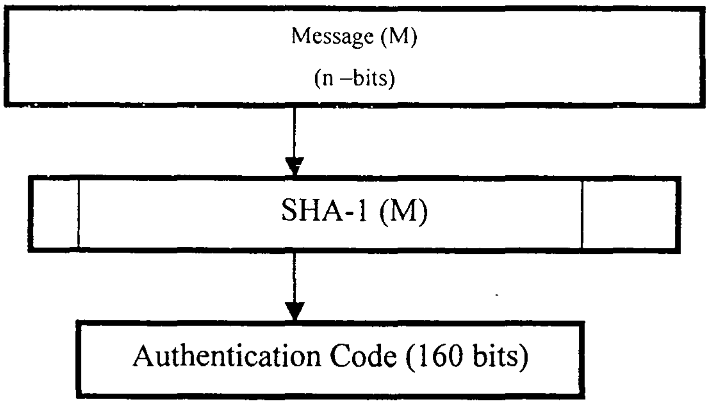


Figure 2.7: The use of SHA-1 hash algorithm

The output from the function then gets appended to the message

$$M \,\|\, \text{SHA-1 } (M)$$

The entire string [M || SHA-1 (M)] then gets transmitted to the intended recipient. The recipient then verifies the received hash value with the received message as input, as illustrated in Figure 2. 7.

Although a lot has been written about the security of SHA-1, it is not secure when used in this way. The reason should be quite obvious to the reader. Both the message and hash value are transmitted in the clear. If an attacker was to intercept the transmission of say, a payment instruction, he could easily alter the instruction for financial gain. All the attacker need s to know is the structure of the message, and the specifics of the hash algorithm used. He could then alter the message, compute a new hash value, and transmit the altered message and the new hash value on the network. To a payments engine, the message would seem legitimate and it would surely process the transaction to the benefit of the attacker. A more secure solution is needed. Is a keyed hash the answer?

2.4.2 Keyed Hash

A keyed hash is defined as a hash function that takes as an input a message of arbitrary length and a key of fixed length and computes the corresponding output string. In this instance we will make use of the SHA-1 hash algorithm and a secret 160-bit key shared between the sender and receiver. Figure 2.8 illustrates this procedure graphically.

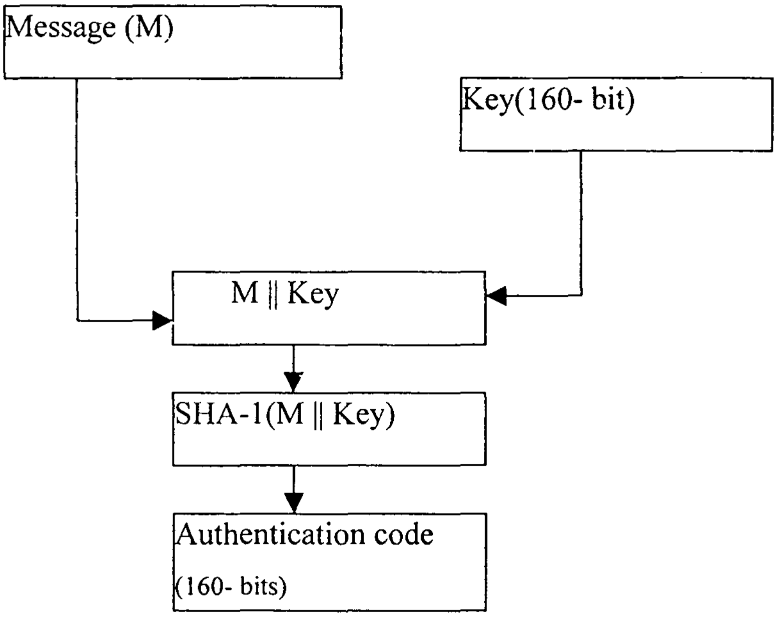


Figure 2.8: Keyed HASH

Although this method of using a hash function makes it more difficult for the attacker to construct a false message during transmission, it is still not cryptographically secure. The key appended to the message and the corresponding authentication code is sent in the clear. An attacker can easily extract the key, construct a new false message, and re-compute the corresponding authentication code. More security is needed to ensure the integrity of data during transmission for a financial transaction. HMAC is a new technology that we explain in the next section.

2.4.3 HMAC

HMAC is used in combination with a cryptographic hash function specified in the Federal Information Processing Standard (FIPS) document [Keyed-Hash Message Authentication Code (HMAC)]. HMAC uses a secret key for the calculation and verification of the MACs.

The main goals behind the HMAC construction are:

- To use available hash functions without modifications; in particular, hash functions that perform well in software, and for which code is freely and widely available,
- To preserve the original performance of the hash function without incurring a significant degradation,
- To use and handle keys in a simple way,
- To have a well-understood cryptographic analysis of the strength of the authentication mechanism based on reasonable assumptions on the under lying hash function, and
- To allow for easy replaceability of the under lying hash function in the event that faster or more secure hash functions are later available.

To determine the HMAC the following steps are to be followed:

1. If the length of K = B, set K_0 = K. Go to step 2. If the length of K > B, hash K to obtain an L byte string: K = H (K). If the length of K < B, append zeros to the end of K to create a B-byte string K_0 (e.g., if K is 20 bytes in length and B = 64, then K will be appended with 44 zero bytes 0x00).
2. Exclusive-Or K_0 with ipad to produce a B-byte string: $K_0 \oplus$ ipad.
3. Append the stream of data 'text' to the string resulting from step 4: ($K_0 \oplus$ ipad) || text.
4. Apply H to the stream generated in step 5: H (($K_0 \oplus$ipad) || text).
5. Exclusive-Or K_0 with opad: $K_0 \oplus$ opad.
6. Append the result from step 6 to step 7: ($K_0 \oplus$ opad) || H (($K_0 \oplus$ipad) || text).
7. Apply H to the result fro m step 8: H (($K_0 \oplus$opad) || H (($K_0 \oplus$ipad) || text)).
8. Select the leftmost t bytes of the result of step 9 as the MAC.

Where:

B Block size (in bytes) of the input to the FIPS-approved hash function;

H FI PS-approved hash function, e.g., FIPS 180-1,

ipad Inner pad; the byte x'36' repeated B times.

K Secret key shared between the originator and the intended receiver(s).

K_0 The key K with zeros appended to form a B byte key.

L Block size (in bytes) of the output of the FIPS-approved hash function;

opad Outer pad; the byte x'5c' repeated B times.

t The number of bytes o f MAC.

text The data on which the HMAC is calculated; the length of the data is n bits, in this case 160 for SHA-1

x'N' Hexadecimal notation, where each 'N' represents 4 binary bits.

|| Concatenation

$\oplus$ Exclusive-Or operation.

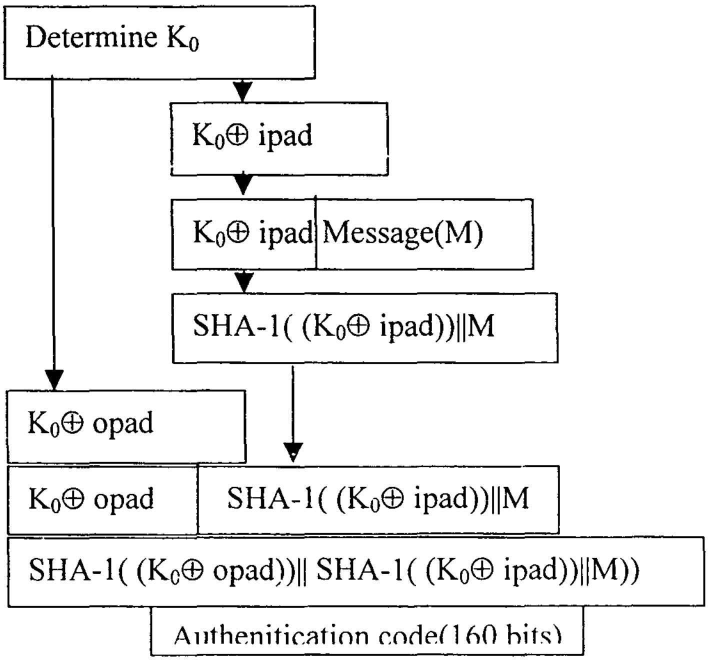

Determine K_0
$K_0\oplus$ ipad
$K_0\oplus$ ipad
Message(M)
SHA-1(($K_0\oplus$ ipad))||M
$K_0\oplus$ opad
$K_0\oplus$ opad
SHA-1(($K_0\oplus$ ipad))||M
SHA-1(($K_0\oplus$ opad))|| SHA-1(($K_0\oplus$ ipad))||M))
Authenitication code(160 bits)

Figure 2.9: HMAC implementation

Although an HMAC is more secure than a hash or a keyed hash, it still does not satisfy our needs. The HMAC might seem complex, and certainly is not easy to decode, it is still not cryptographically strong. SHA-1 does not provide confidentiality, and the exclusive-OR function is certainly not difficult to replicate. Thus attackers can reproduce an HMAC given the clear text message and the authentication code, as the keys can be extracted with relative ease. We need a mechanism that can ensure that an attacker cannot derive the authentication code from the transmission of the message and the appended authentication code. The next section will look at such a mechanism.

2.4.4 Message Authentication Code (MAC)

In this section we will look a SHA-1 Triple DES MAC in CBC-mode. DES as we know, stands for the Data Encryption Standard. The CBC-mode points to the mode of operation in which we would like to implement the triple DES function, Cyclic Block Chaining in this instance. In computing the authentication code in this method we follow the following simple steps:

1. Compute the SHA-1 hash value for the message being transmitted.
2. Encrypt the SHA-1 hash value using Triple DES in CBC-mode using two secret keys.
3. Take the last 8-byte block (CN) output from the Triple DES function as the MAC.
4. Append the MAC to the message and transmit to the recipient.

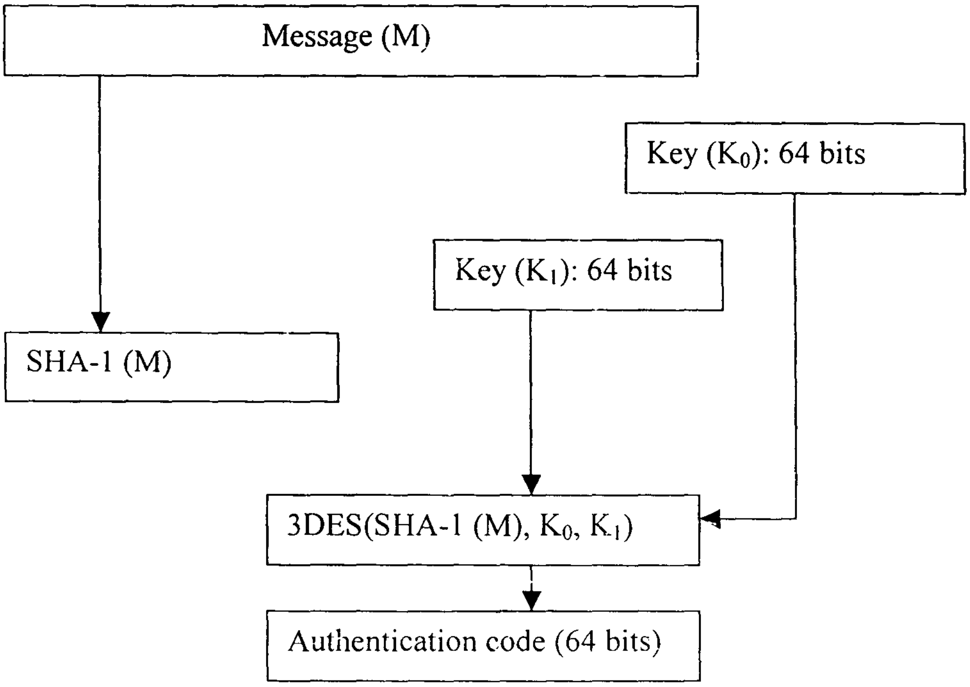


Figure 2.10: SHA- 1 3DES MAC

I he authentication code computed in this fashion can be regarded as cryptographically secure. Even when an attacker intercepts the clear text message and the MAC, he cannot extract the keys from the MAC. This implies that the attacker cannot alter the message and re-compute the MAC, as he has no access to the keys needed. Another reason for computing the MAC in this fashion is that industrial Hardware Security Modules (HSM) can be used to perform this function.

In a typical industrial application, the SHA-1 will be computed in software, the SHA-1 value is then sent to a HSM to compute the MAC on this value. The reason for doing the computation this way is due to the nature of HSMs. A typical HSM is built for transactional encryption techniques, like PIN verification. If we now send a big file to the HSM to MAC in the normal 3DES CBC-mode, the HSM will take a long time to compute all the blocks required. While this MAC operation is taking place on the HSM, no other transactions can be verified or completed. In a high volume transactional environment, this can be catastrophic. By computing the SHA- 1 value in software, we restrict the size of the data sent to the HSM for encryption, thus speeding up the computation of the MAC.

Although a MAC computed in this fashion is cryptographically secure, it does not give us non-repudiation. In the next section we will look at the RSA digital signature that will provide this service to us as well.

2.4.5Digital Signatures

Digital signatures can only be accomplished by means of public key cryptography. Computing the hash of a message and encrypting that hash with the sender's private key achieves this. When the receiver receives the message, he decrypts the encrypted hash with the sender's public key, and verifies the received hash with the hash he himself computes from the received message. In this way he can be sure that the person that sent him the message is the authenticated sender, and that the message has not been tampered with during transmission. Figure 2. 11 illustrate this graphically.

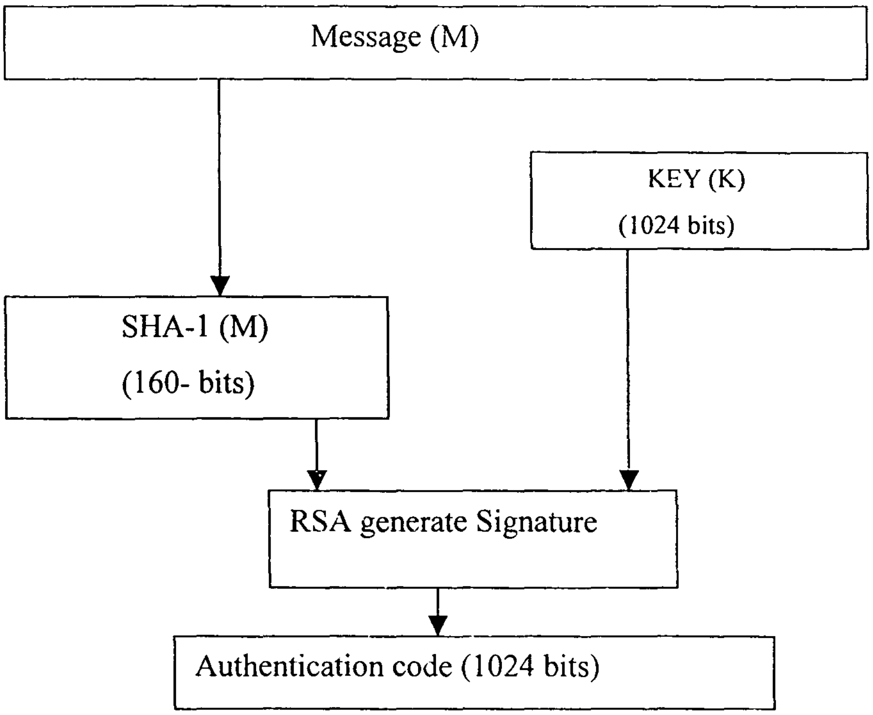


Figure 2.11: Generation of a RSA Digital Signature

In the industry, people are always transacting over untrusted networks. In order for an organization to be able to communicate with other organizations or individuals, it is very important that they adhere to international standards. These standards assure that all communicating parties are using the same protocols to communicate. Just as TCP/IP became an international standard, so too have several cryptographic functions.

In the previous section we made use of SHA-1, 3DES and RSA as examples of these kinds of technologies, because of their standing as inter national standards. The author does not hereby imply that they are the only, or even the best tools to make use of. The selection of these functions comes from experience with communication failure due to the communicating parties making use of non-standards based options.

The public algorithms mentioned above have been scrutinized by subject matter experts. Their inherent vulnerabilities are therefore well known and documentea. This is not the case with proprietary algorithms. Areas of vulnerability within proprietary algorithms are unknown and could pose serious threats. In a highly volatile environment such as banking, time-to-market is of utmost importance. Time and costs involved with programming proprietary algorithms may lead to a window of opportunity being missed.

Table 2.2 summarizes the services provided by each of the algorithms described above.

Crypto Tool	Authentication	Integrity	Non-Repudiation	Confidentiality
HASH	0	1	0	0
Keyed HASH	1	2	0	0
HMAC	2	3	1	0
MAC	3	5	3	0
RSA signature	5	5	5	0

Table 2-2: Services provided by cryptographic authentication

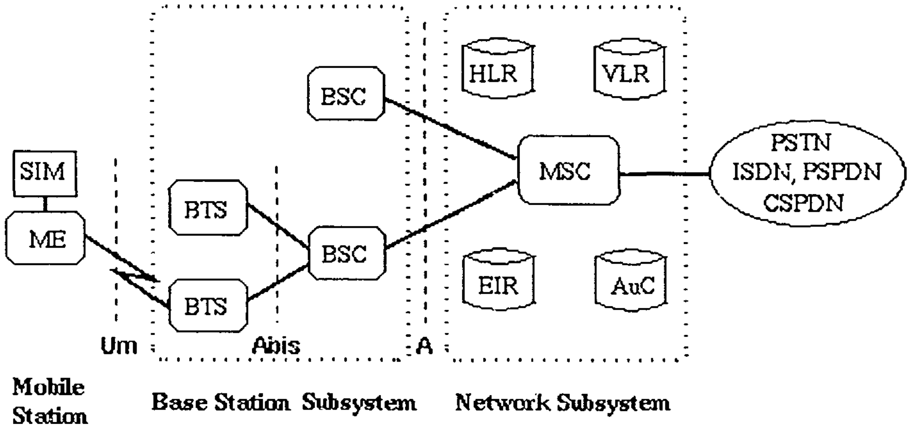
Where: 0 = No service provided, 5 = Full service provided.

Chapter 3

SECURITY IN GSM SYSTEM

3.1Architecture of the GSM network

A GSM network is composed of several functional entities, whose functions and interfaces are specified. Figure3.1 shows the layout of a generic GSM network. The GSM network can be divided into three broad parts. The subscriber carries the Mobile Station. The Base Station Subsystem controls the radio link with the Mobile Station. The Network Subsystem, the main part of which is the Mobile services Switching Center (MSC), performs the switching of calls between the mobile users, and between mobile and fixed network users. The MSC also handles the mobility management operations. Not shown is the Operations and Maintenance Center, which oversees the proper operation and setup of the network. The Mobile Station and the Base Station Subsystem communicate across the Um interface, also known as the air interface or radio link. The Base Station Subsystem communicates with the Mobile services Switching Center across the A interface.



SIM	Subscriber Identity Module	BSC	Base Station Controller	MSC	Mobile services Switching Center
ME	Mobile Equipment	HLR	Home Location Register	EIR	Equipment Identity Register
BTS	Base Transceiver Station	VLR	Visitor Location Register	AuC	Authentication Center

Figure 3 . 1 General architecture of a GSM network

3.1.1Mobile Station

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

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

3.1.2Base Station Subsystem

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

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

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

3.1.3Network Subsystem

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

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

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

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

3.2Security features offered by GSM

GSM specification 02.09 identifies three areas of security that are addressed by GSM.

- **Authentication of a user-** this deals with the ability for a mobile phone to prove that it has access to a particular account with the operator
- **Data and signaling confidentiality**-this requires that all signaling and user data are protected against interception by means of ciphering
- **Confidentiality of a user**-this deals with the fact that when the network needs to address a particular subscriber, or during the authentication process, the unique IMSI(international mobile subscriber identity) should not be disclosed in plaintext (unciphered). This means someone intercepting communications should not be able to learn if a particular mobile user is in the area

3.2.1Authentication

Authentication is needed in a cellular system to prohibit an unauthorized user from logging into the network claiming to be a mobile subscriber. If this were possible, it would be easily possible to "hijack" someone's account and impersonate that person (or simply making that person pay for the services). In fact, this was possible in some earlier cellular systems.

In order to solve this problem, some sort of challenge needs to be issued by the network, which the mobile phone (MS) must respond to correctly.

3.2.1.1 The SIM card

Many users of GSM will be familiar with the SIM (Subscriber Identity Module) – the small smart card that is inserted into a GSM phone. On its own, the phone has no association with any particular network. The appropriate account with a network is selected by inserting the SIM into the phone. Therefore the SIM card contains all of the details necessary to obtain access to a particular account. These details come down to just 2 items of information.

- The IMSI – International Mobile Subscriber Identity – a unique number for every subscriber in the world. It includes information about the home network of the subscriber and the country of issue. This information can be read from the SIM provided there is local access to the SIM (normally protected by a simple PIN code). The IMSI is a sequence of up to 15 decimal digits, the first 5 or 6 of which specify the network and country (i.e. 50501 for Telstra, Australia)

- The K_i – the root encryption key. This is a randomly generated 128-bit number allocated to a particular subscriber that seeds the generation of all keys and challenges used in the GSM system. The K_i is highly protected, and is only known in the SIM and the network's AuC (Authentication Centre). The phone itself never learns of the K_i, and simply feeds the SIM the information it needs to know to perform the authentication or generate ciphering keys. **Authentication and key generation is performed in the SIM**, which is possible because the SIM is an intelligent device with a microprocessor.

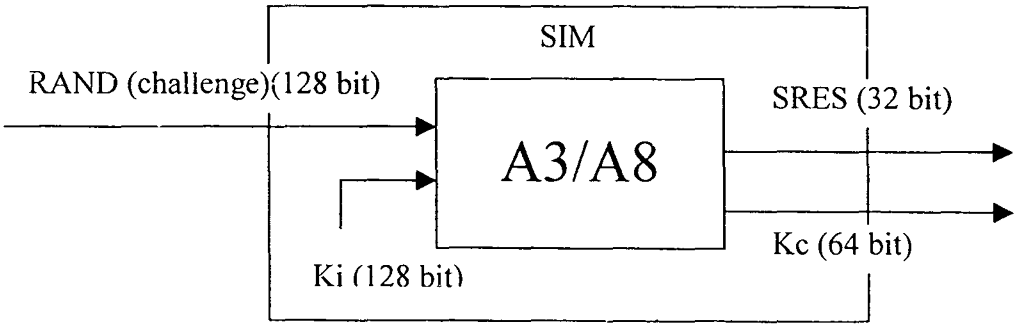


Figure 3.2 The SIM Authentication Concept

3.2.1.2 The A3 algorithm and authentication procedure

Now that we have established that there is a 'secret' K_i known only in the SIM and the network, the authentication procedure simply has to involve the SIM (via the phone) proving knowledge of the K_i. Of course, we could simply submit the K_i to the network for comparison when the network asks for it, but this is highly insecure, since the K_i

could be intercepted. Instead, the network generates a 128-bit random number, known as the RAND, which it then uses the A3 algorithm (see figure) to mathematically generate an authentication token known as the SRES. It then sends the RAND to the phone for the phone to do the same. The SIM generates the 32-bit SRES, which is returned to the network for comparison. If the received SRES matches the network's generated SRES, then the K_i's must be the same (to a high mathematical probability), and the phone has proved knowledge of the K_i and is thus authenticated.

The RAND must obviously be different every time. Otherwise, if it were the same, an attacker could impersonate the user by sending the same SRES.

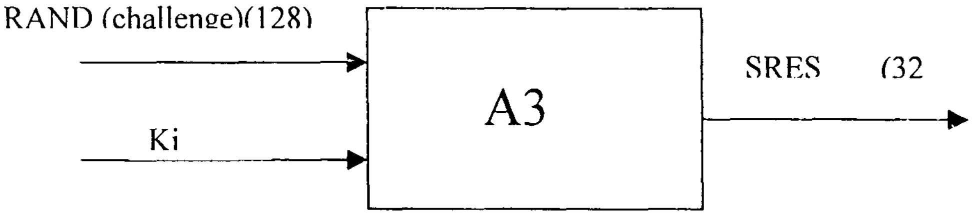


Figure 3.3

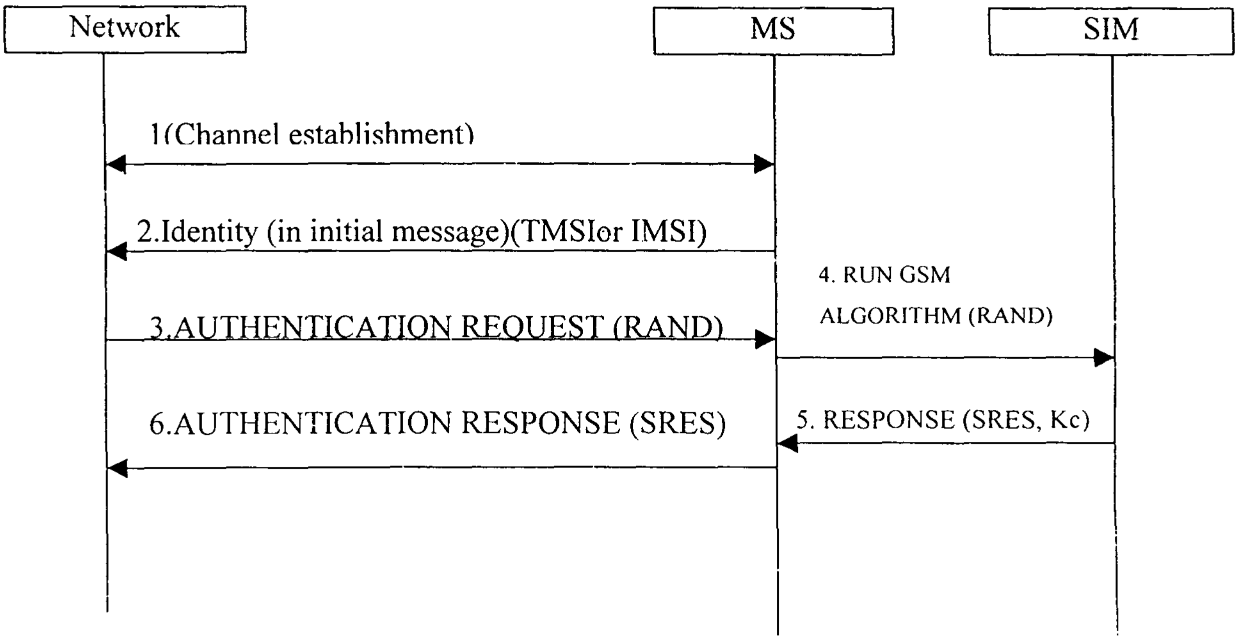

Network
MS
SIM
1(Channel establishment)
2.Identity (in initial message)(TMSIor IMSI)
3.AUTHENTICATION REQUEST (RAND)
4. RUN GSM ALGORITHM (RAND)
5. RESPONSE (SRES, Kc)
6.AUTHENTICATION RESPONSE (SRES)

Figure 3.4

The above procedure can be summarized as follows:

(prior) – the network pre-generates a RAND and pre-calculates the SRES for that subscriber.

1 Some connection is attempted between the phone and the network.

2.The phone submits its identity. All potential messages used at the start of connection contain an identity field. Where possible, it avoids sending its IMSI in plaintext (to prevent eavesdroppers knowing the particular subscriber is attempting a connection). Instead, it uses its TMSI (Temporary Mobile Subscriber Identity).

3.The network sends the AUTHENTICATION REQUEST message containing the RAND.

4.The phone receives the RAND, and passes it to the SIM, in the RUN GSM ALGORITHM command.

5.The SIM runs the A3 algorithm, and returns the SRES to the phone.

6.The phone transmits the SRES to the network in the AUTHENTICATION RESPONSE message.

7.The network compares the SRES with its own SRES. If they match, the transaction may proceed. Otherwise, the network either decides to repeat the authentication procedure with IMSI if the TMSI was used, or returns an AUTHENTICATION REJECT message.

3.2.1.3 Authentication failure

If authentication fails the first time, and the TMSI was used, the network may choose to repeat the authentication with the IMSI. If that fails, the network releases the radio connection and the mobile should consider that SIM to be invalid (until switch-off or the SIM is re-inserted).

3.2.1.4 More on the A3 algorithm

The A3 algorithm does not refer to a particular algorithm, rather the algorithm the operator has chosen to be implemented for authentication. The most common implementations for A3 are COMP128v1 and COMP128v2. In fact, both of these algorithms perform the function of both A3 and A8 (the ciphering key generation algorithm – discussed later) in the same stage. Whenever the SIM is asked to compute the SRES (with the RUN GSM ALGORITHM command) it also computes a new K_c (ciphering key – discussed later). Thus not only is the authentication procedure used to verify a user, it is also used whenever the network wishes to change keys.

3.2.2Ciphering

Ciphering is highly important to protect user data and signaling data from interception. The GSM system uses symmetric cryptography - the data is encrypted using an algorithm which is 'seeded' by the ciphering key – the K_c. This same K_c is needed by the decryption algorithm to decrypt the data. The idea is that the K_c should only be known by the phone and the network. If this is the case, the data is meaningless to anyone intercepting it.

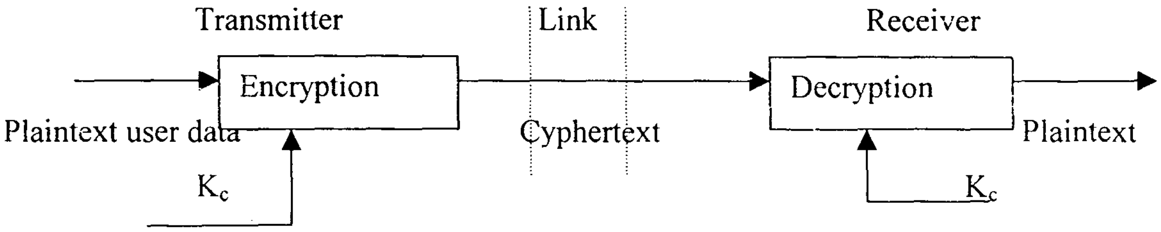


Figure 3.5

The K_c should also frequently change, in case it is eventually compromised. The method of distributing the K_c to the phone is closely tied in with the authentication procedure discussed above.

Whenever the A3 algorithm is run (to generate SRES), the A8 algorithm is run as well (in fact the SIM runs both at the same time). The A8 algorithm uses the RAND and K_i as input to generate a 64-bit ciphering key, the K_c, which is then stored in the SIM and readable by the phone. The network also generates the K_c and distributes it to the base station (BTS) handling the connection.

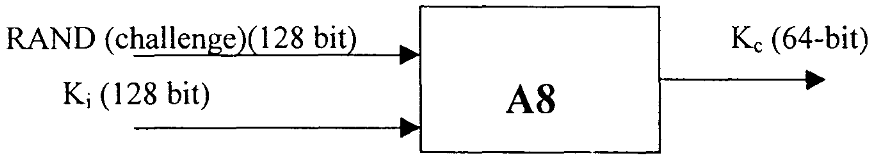


Figure 3.6

At any time, the network can then order the phone to start ciphering the data (once authenticated) using the K_c generated. The network can pick from a number of algorithms to use, as long as the phone supports the one chosen (this is indicated to the network earlier in a classmark message, which specifies the phone's capabilities). The ciphering algorithm works by generating a stream of binary data (the cipher block), which is modulo-2 added (XORed) with the user data, to produce the ciphered text that is transmitted over the air. The data is decrypted by XORing the received data with the cipher block, which should be the same if the K_c is the same.

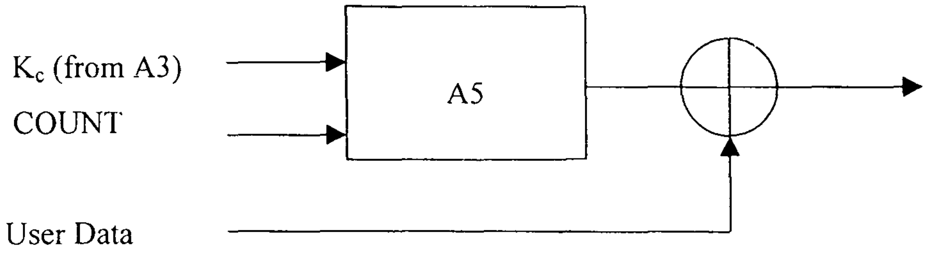


Figure 3.7

The algorithm is also 'seeded' by the value COUNT, which is based on the TDMA frame number, sequentially applied to each 4.615ms GSM frame. Internal state of the algorithm is flushed after each burst (consisting of 2 blocks of 57 bits each). In the case of multislot configurations, different cipher contexts are maintained for each timeslot. The same base

K_c is used, however it is manipulated for each timeslot by XORing bits 32-34 of the K_c with the 3-bit timeslot number (0-7).

3.2.2.1 Ciphering algorithms

As mentioned above, the network can choose from up to 7 different ciphering algorithms (or no ciphering), however it must choose an algorithm the phone indicates it supports.

Currently there are 3 algorithms defined – A5/1, A5/2 and A5/3. A5/1 and A5/2 were the original algorithms defined by the GSM standard and are based on simple clock controlled LFSRs. A5/2 was a deliberate weakening of the algorithm for certain export regions, where A5/1 is used in countries like the US, UK and Australia. A5/3 was added in 2002 and is based on the open Kasumi algorithm defined by 3GPP.

3.2.3Anonymity

As mentioned above, one of the main goals of GSM security was to avoid having to use the IMSI (International Mobile Subscriber Identity) in plaintext over the radio link. This is avoided by addressing the phones by a 32-bit TMSI (Temporary Mobile Subscriber Identity), which is only valid in a particular Location Area (i.e. one paging domain). The subscriber addresses it or is paged by the 32-bit TMSI from then on.

The TMSI is updated at least during every location update procedure (i.e. when the phone changes LA or after a set period of time). The TMSI can also be changed at any time by the network. The new TMSI is sent in ciphered mode whenever possible so an attacker cannot maintain a mapping between an old TMSI and a new one and "follow" a TMSI.

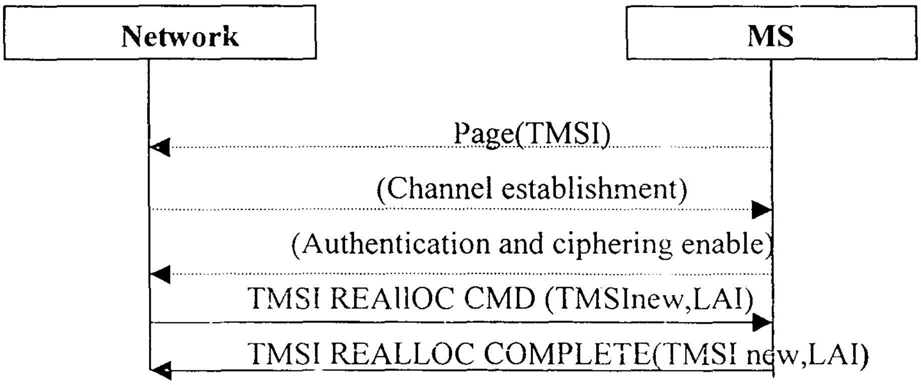

Network
MS
Page(TMSI)
(Channel establishment)
(Authentication and ciphering enable)
TMSI REAllOC CMD (TMSInew,LAI)
TMSI REALLOC COMPLETE(TMSI new,LAI)

Figure 3.8 :Allocating a new TMSI (when not location updating or location change during connection)

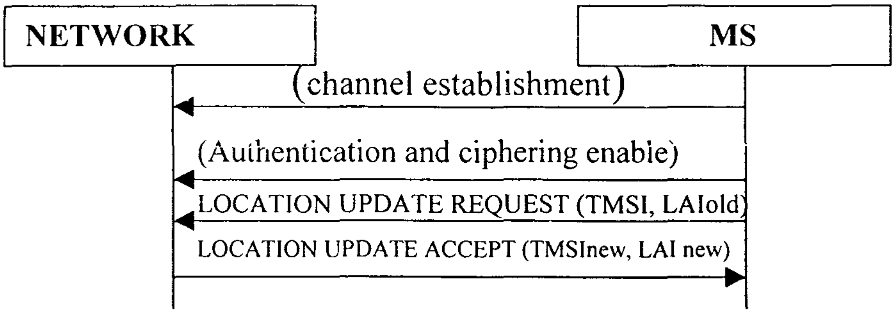

NETWORK
MS
(channel establishment)
(Authentication and ciphering enable)
LOCATION UPDATE REQUEST (TMSI, LAIold)
LOCATION UPDATE ACCEPT (TMSInew, LAI new)

Figure 3.9 :Allocating a new **TMSI** (when moving into a new location area in idle mode)

The phone must store the TMSI in non-volatile memory (so it is not lost at switch off). It is normally stored in the SIM. Initially of course the phone will have no TMSI, and thus is addressed by its IMSI. Once ciphering has commenced the TMSI is allocated. The VLR controlling the LA in which the TMSI is valid maintains a mapping between the TMSI and IMSI such as that the new VLR (if the MS moves into a new VLR area) can ask the old VLR who the TMSI (which is not valid in the new VLR) belonged to.

3.2.4Implementations of A3, A8[15]

Although the design of the GSM system allows an operator to choose any algorithm they like for A3 & A8, many decided on the one that was developed in secret by the GSM association, COMP128.

COMP128 eventually ended up in public knowledge due to a combination of reverse engineering and leaked documents, and serious flaws were discovered (as discussed below).

Some GSM operators have moved to a newer A3/A8 implementation, COMP128-2, a completely new algorithm which was also developed in secret. This algorithm for now seems to have addressed the faults of the COMP128 algorithm, although since it has yet to come under public scrutiny it may potentially be discovered via reverse-engineering and any possible flaws could be learned.

Finally, the COMP128-3 algorithm can also be used, it is simply the COMP128-2 algorithm, however all 64-bits of the K_c are generated, allowing maximal possible strength from the A5 ciphering algorithm (COMP128-2 still sets the 10 rightmost bits of the K_c to 0), deliberately weakening the A5 ciphering.

3.3Introduction to Security Flaws

The GSM infrastructure for subscriber authentication and confidentiality of communication sessions represented a major advance over first generation analogue cellular system. However as GSM has matured and expanded its reach across Europe and beyond, its basic security mechanism have come under increasing attacks and criticism. Given the strong belief in the security community that only protocols that can be tested should be trusted (that security should depend on the secrecy of keys and not of algorithms), it was inevitable that GSM would be attacked for its dependency on the proprietary A3, A8, and A5 algorithms. Many security analysts view these algorithms as cryptographically weak and subject to intrusion by government agencies, in addition to well-equipped hackers. In this chapter we will look at flaws with GSM security in order to understand what more needs to be done to make mobile commerce secure.

3.4 Problems With GSM Security

3.4.1Common implementation of A3/A8 is flawed

GSM makes use of the A3 and A8 algorithms to authenticate the user and generate the session key for secure transmission of user and signaling data over the air. GSM service providers typically implement both A3 and A8 with an algorithm called COMP128, which generates the 64-bit K_c and the 32-bit SRES from the 128-bit RAND and the 128–bit K_i input

COMP128 was reverse engineered at Berkley in 1998, and cryptanalysis by Berkley researchers indicates that the protocol can be broken with 219 queries from rogue base station to the GSM SIM card, which can be achieved in eight hours. Analysis of the GSM application of COMP128 further revealed that it had apparently been deliberately weakened. The algorithm calls for a 64-bit key, but of this total, ten key bits are consistently set to zero, dramatically reducing the security offered by the A8 implementation.

3.4.2Network does not authenticate itself to a phone

Under the GSM authentication protocol, the GSM base station authenticates the mobile station, which seeks to establish a communication session, i.e. unilateral authentication. However, the opposite is not true. Thus the mobile station has no guarantee that it is not communicating with a node, which is impersonating a GSM base station. To make the situation worse, the same random challenge (RAND) that is used to authenticate mobile station, when presented to the A8 algorithm also becomes the seed for the generation of a session key K_i. Furthermore, the authentication challenge response message protocol does not include a time stamp. Thus if a rogue station does successfully impersonate a GSM base station, it may secure a session key that will allow the decryption of any message sent with the same key over a potentially prolonged period.

3.4.3Vulnerabilities in the subscriber identity confidentiality mechanism

GSM specifications have gone to a great length to avoid phones being addressed (i.e. paged) or identifying themselves in plaintext by their IMSI. This is supposed to prevent an eavesdropper listening in on the initial plaintext stage of the radio communication learning that a particular subscriber is in the area. Thus where possible the network pages by their TMSI (Temporary Mobile Subscriber Identity) and maintains a database in the VLR mapping TMSIs to IMSIs.

If the network somehow loses track of a particular TMSI, and therefore cannot determine who the user is, it must then ask the subscriber its IMSI over the radio link, using the IDENTITY REQUEST and IDENTITY RESPONSE mechanism. Obviously, the connection cannot be ciphered if the network does not know the identity of the user, and thus the IMSI is sent in plaintext.

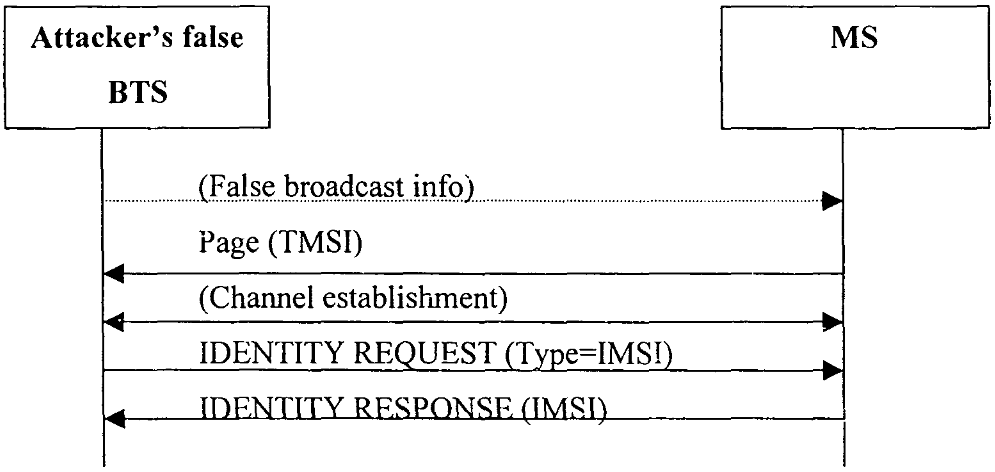


Figure 3.9

Combined with the previously described flaw that the network does not have to authenticate itself to a phone, an attacker can use this to map a TMSI to its IMSI. The attacker accomplishes this by imitating a legitimate base station of the subscriber's network and paging that subscriber by its IMSI .The subscribers phone will then establish a radio connection, and then the attacker can simply send the subscriber the IDENTITY REQUEST (Identity Type=IMSI) message, and the phone will respond with IMSI.

3.4.4Ciphering occurs after FEC

In the GSM system, like all digital communications system, forward error correction (FEC) is used over the radio link to assist in the correction from errors caused by noise or signal fading. FEC works by adding redundancy to the data stream, thus increasing the amount of bits to transfer.

The problem in GSM is that ciphering occurs after FEC, meaning the redundant stream of bits is then modulo-2 (XORed) added with the ciphering stream, meaning the known redundancy patterns could be used to assist in a crypt –analytical attack. In the most simple example suppose the error correction codes consist of the convolution code polynomials G_1=1 and G_2 =2,corresponding to a simple twice repletion of bits. Suppose a received 6-bit sequence is 101110. Thus, it can be concluded that the 6-bit ciphering stream consisting of the bits abcdef is such that:

a=b`, c=d, and e=f`

Of course it is far more complicated than that; in GSM the data is interleaved over many blocks and the coding on most encrypted channels is a ½ rate (i.e. non punctured) convolution code generated by the polynomial with m from 5-7.Furthermore, on voice channels only certain bits are protected with the convolution code (unequal error protection) However combined with knowledge of the A5 ciphering algorithm, a cryptanalyst could use this information to reduce the number of keys to search from the ideal(2^{64}) to something less.

3.4.5Flaws in A5/1 and A5/2 algorithm[16]

The A5/1 output is based on the modulo-2 summed output of 3 LFSRs whose clock inputs are controlled by a majority function of certain bits in each LFSR.

However, Alex Biryukov, Adi Shammir and David Wagner demonstrated in a paper that A5/1 could be cracked in 1 second on a typical PC (however large precomputed tables are required, amounting to about 2^{36} bytes or 64G).

The attack exploits flaws in the algorithm when storing theses tables utilizing a combination of what has been learnt through statistical analysis of the states the algorithm steps through, as well as exploiting the poor single-bit taps used to control the LFSR clocks.

A5/2 is a deliberately weakened version of A5/1,which has been demonstrated to be also flawed .A5/2 can be cracked on the order of about 2^{16}, and thus is even weaker than A5/1.

As can be seen from above discussion, there are numerous flaws in both GSM authentication and encryption algorithm. This is largely due to the fact that the algorithms were developed in isolation and not put under public security. Even though the GSM forum tried to keep the algorithm secret, some of them were reverse engineered or even leaked to the public. Whenever security is obtained through obscurity, failure is guaranteed.

Chapter 4

SECURE M- COMMERCE BASED ON WIG

4.1 Introduction

Wireless Internet Gateway (WIG) is a technology available on the GSM network that enables a network operator to effect wireless GSM based payment instructions to financial institutions. In this chapter the architecture of the WIG application, dependencies on other technologies, and the security thereof as implemented will be discussed.

We will firstly look at the security of the service technologies that can provide WIG applications to us, namely Unstructured Supplementary Services Data (USSD) and Short Message Service (SMS). Thereafter we will look at enabling applications by means of the SIM Application Toolkit (SAT) and the security of the SAT.

Once the whole architecture is built we will do a brief analysis of the WIG solution and then propose a model that will provide secure m-Commerce transactions via WIG.

4.2 Unstructured Supplementary Services Data (USSD)

4.2.1 Introduction to USSD

USSD is a session-oriented technology, unlike SMS, which is a store-and-forward technology. Turnaround response times for interactive applications are shorter for USSD than SMS because of the session-based feature of USSD [9].

Users do not need to access any particular handset menu to access services with USSD, as they can enter the USSD command direct from the initial mobile phone screen. USSD commands are routed back to the home mobile network's HLR, allowing the ability for services, based on USSD, to work just as well and in exactly the same way when users are roaming. USSD works on all existing GSM mobile phones, and both SIM Application Toolkit and Wireless Application Protocol (WAP) support USSD [9]

4.2.2 Operation of USSD

In operation, USSD is used to send text between the user and an application. USSD should be thought of as a trigger rather than an application itself, as it enables other applications such as prepaid top-ups. USSD provides an ideal way for subscribers to request changes to their class of service, request that enhanced services are per formed, or to perform a mobile originated payment instruction. To achieve this, the sequence of operations is as follows:

- A Subscriber sends a mobile originating USSD message.
- The USSD message is routed to the subscriber's HLR in accordance with the GSM recommendations.
- The HLR forwards the USSD message to the USSD Gateway.
- The USSD Gateway communicates the message to external applications using TCP/IP, which is more convenient for integration with commercial computing platforms.
- The external system interprets the message and performs the action indicated by the content of the message.
- Within a time-out period, the external system acknowledges successful receipt of the message to the mobile via the USSD Gateway. The external system can later asynchronously send further information to the mobile as a Short Message via an SMS.

4.2.3 Benefits offered by USSD

The primary benefit of USSD is that it allows for very fast communication between the user and an application. Most of the applications enabled by USSD are menu based and include services such as mobile prepay and chat. Some key benefits of USSD are [9]:

- Easy to use: Keying a digit string can be easier for a user than formatting a short message. Strings may be stored under abbreviated dial keys on the handset.

- USSD messages are very flexible in both length and content.
- USSD is faster than SMS.
- Roaming is supported. Because messages are exchanged with your HLR, services are still available when roaming.
- Service access codes and service names may be downloaded to the handset using Over the Air Programming. This makes it even easier for the user to get started.

4.2.4 Difference between USSD and SMS

The question might arise as to what are the differences between USSD and SMS. USSD is not store and forward and does not offer retries. This enables it to be simpler and faster than SMS. USSD does not offer guaranteed delivery, but any failures are reported back to the originator. USSD should achieve many times the speed of SMS due to its simplicity and reduced reliance on non-volatile storage. USSD offers a simple TCP/IP interface to external applications, which knows nothing of the SS7 network. Routing to applications is achieved via a simple service code, which is contained in the USSD message. The interpretation of the Service Code is achieved by configuration of the USSD Gateway and by the actions of the external application to which the service code relates. The external applications can be on any machine reachable by a TCP/IP network [9].

4.2.5 Security of USSD

USSD cannot be viewed as a trusted or secure channel from a financial point of view. To understand this concept, we need to look at the structure of a USSD message. The Formatting of a USSD messages can be summarized as follows:

- An asterisk is used to separate each of the parameters.
- A service code of 2 or 3 digits is entered.
- Supplementary information can then be entered. This may be of variable length. As an example, a PIN may be used as a measure of security.
- The # key terminates a request.

This message string is then transmitted across the network to the subscribers HLR. We know that the transmission is not encrypted or secured with an integrity check on the GSM backbone. This makes these USSD messages highly vulnerable to attack. An attacker can thus easily delete, alter, or even fabricate false messages on the network. For service alteration requests this does not pose a problem. If however we are initiating financial transactions on this channel, the picture changes rapidly.

Let's look at a possible attack on a financial application in order to demonstrate the risk associated with the USSD channel.

A client of Bank A is resident in INDIA, but is currently roaming in US. He has registered with Bank A as an m-Banking user via their normal client registration guidelines. Bank A provides him a payment application, based on USSD, whereby he can pay money from his Credit Card to any other, Card Association endorsed, Credit Card. Client A just conducted a business deal and needs to pay Client B, based at Bank B, the amount specified by the transaction. Client A decides to make use of his mobile facility to effect the transaction and retrieves the relevant information from the intended recipient.

Client A now initiates the payment instruction via USSD to Bank A as depicted in Figure 4.1. Client A has memorized the instruction, and also needs to enter his PIN for the transaction to be affected. Client A enters the following string into his GSM based mobile phone: *184*1234*1*50000*5120123412341234*+411234567890#, where:
*184 = the service co de *1234 = PIN of Client A
*1 = the account indicator from which the transaction should take place
*50000 = the amount to be paid in Client A's local currency
*5120123412341234 = the account number that the payment should be made too.
*+411234567890 = the mobile number where the payment confirmation should be sent.
# = Request terminator.

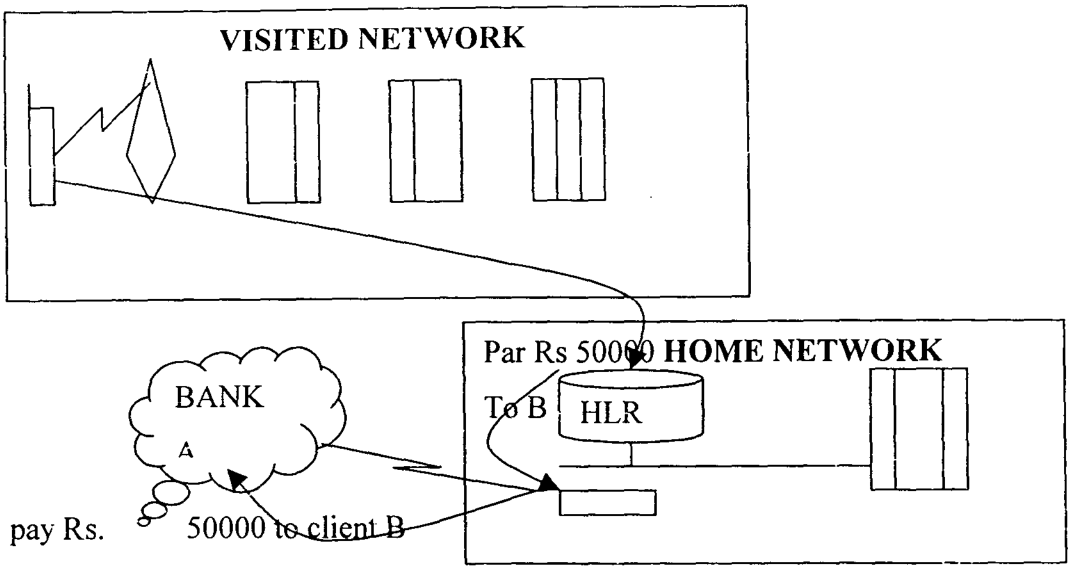
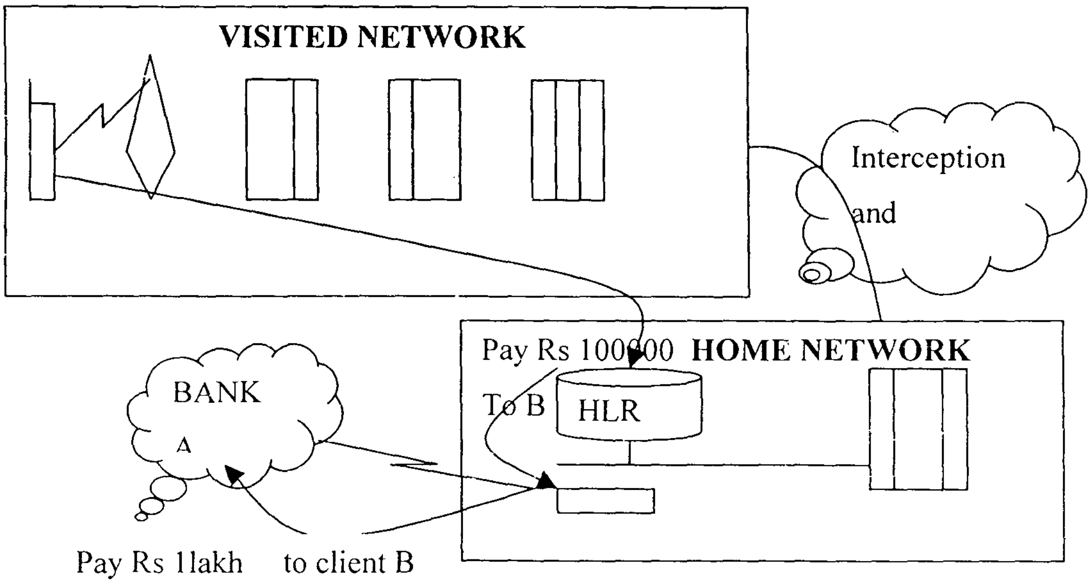


Figure 4.1: Normal USSD Transaction Flow

Client B also has some friends in a fraud syndicate and tells them that Client A will be paying the agreed amount via his m-Banking application within the near future, and that he would like them to steal as much money as possible.

The syndicate in turn has access to the required technology to intercept the transaction and alter it. Due to the fact that there is no encryption or integrity checking on the message during transmission through the GSM backbone, this attack is possible. The attackers can easily alter the amount to be paid, or even the account to which the money is to be paid. Figure 4.2 depicts this attack.



From the above scenario it is evident that some measure of encryption or message integrity checking is required in order to provide a reasonably secure USSD based payment application. USSD cannot provide this service on its own. Another application or technology is required in order to secure this channel.

4.3 Short Message Service (SMS)

4.3.1 Introduction to SMS

SMS appeared on the wireless scene in 1991 in Europe and the GSM standards included short messaging services from the outset. SMS provides a mechanism for transmitting short messages to and from wireless handsets. The service makes use of a short message service center (SMSC), which acts as a store-and-forward system for short messages. The wireless network provides for the transport of short messages between the SMSCs and wireless handsets. In contrast to existing text message transmission services such as alphanumeric paging, the service elements are designed to provide guaranteed delivery of text messages to the destination [10].

A distinguishing characteristic of the service is that an active mobile handset is able to Receive or submit a short message at any time, independent of whether or not a voice or data call is in progress. SMS also guarantees delivery of the short message by the network. Temporary failures are identified, and the short message is stored in the network until the destination becomes available [10].

Out-of-band packet delivery and low-bandwidth message transfer characterize SMS. Initial applications of SMS focused on eliminating alphanumeric pagers by permitting two-way general-purpose messaging and notification services, primarily for voice mail. As technology and networks matured, a variety of services were introduced, including electronic mail and fax integration, paging integration, interactive banking, and information services such as stock quotes [10].

4.3.2 SMS Architecture

Figure 5.3 graphically illustrates the SMS architecture as described in the GSM specifications. A brief description of each element in the SMS architecture follows.

4.3.2.1 Short Messaging Entities

Short messaging entity (SME) is an entity that may receive or send short messages. The SME may be located in the fixed network, a mobile station, or another service center [10].

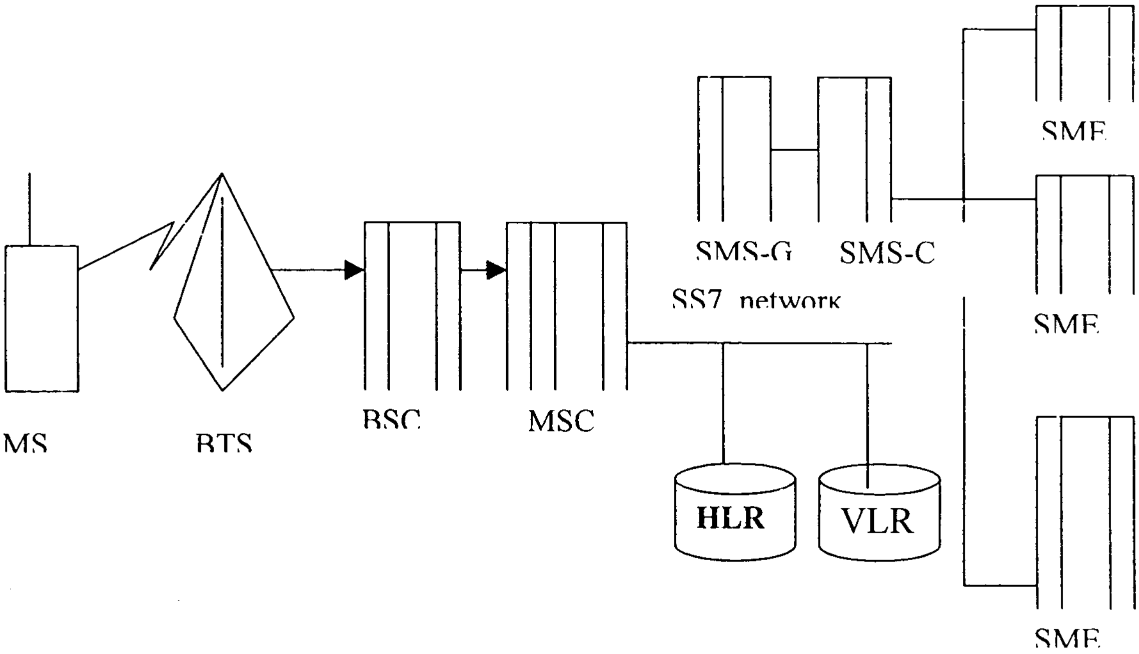


Figure 4.3: The SMS Architecture

4.3.2.2 Short Message Service Center

Short Message Service Center (SMS-C) is responsible for relaying, storing and forwarding of a short message between an SME and mobile station [10].

4.3.2.3 SMS-Gateway/ Interworking Mobile Switching Center

The SMS Gateway Mobile Switching Center (SMS-GMSC) is an MSC capable of receiving a short message from an SMSC, interrogating a HLR for routing information, and delivering the short message to the visited MSC of the recipient mobile station. The

SMS Interworking MSC (SMS–I WMSC) is an MSC capable of receiving a short message from the mobile network and submitting it to the appropriate SMSC. The SMS–GMSC/ SMS–IWMSC are typically integrated with the SMS-C [10].

4.3.2.4 Home Location Register

Upon interrogation by the SMS-C, the HLR provides the routing information for the indicated subscriber. The HLR also keeps previously initiated unsuccessful short message delivery attempts to a specific mobile station. If a previously unreachable mobile station is now recognized by the mobile network to be accessible, that HLR informs the SMS-C of the fact so that the SMS-C can retry the delivery of the undelivered SMSs [10].

4.3.3 SMS Operation

4.3.3.1 SMS Signaling

The Mobile Application Part (MAP) layer defines the operations necessary to support SMS [11]. Both American and international standards bodies have defined a MAP layer using the services of the SS7 transaction capabilities part. The following basic MAP operations are necessary to provide the end-to-end short message service:

- Routing information request [10]:

Before attempting short message delivery, the SMS-C must retrieve routing information to determine the serving MSC for the mobile station at the time of the delivery attempt. This is done by way of an interrogation of the HLR, which is accomplished via the use of the sendRoutingInfoForShortMsg mechanism.

- Point-to-point short message delivery [11]:

The mechanism provides a means for the SMS-C to transfer a short message to the MSC that serves the addressed mobile station and attempts to deliver a message to an
MS whenever the MS is registered, even when the MS is engaged in a voice or data call. The short message delivery operation provides a confirmed delivery service. The operation works in tandem with the base-station subsystem while the message is being

forwarded from the MSC to the MS. The outcome of the operation thus comprises either successful delivery to the mobile, or failure caused by one of several possible reasons. The point-to-point short message delivery is accomplished via the use of the forwardShortMessage mechanism.

- Short message waiting indication [11]:

The operation is activated when a short message delivery attempt by the SMS-C fails due to a temporary failure and provides a means for the SMS-C to request the HLR to add an SMS-C address to the list of SMS-Cs to be informed when the indicated mobile station becomes accessible. This short-message waiting indication is realized via the use of the set message waiting data mechanism.

- Service center alert [11]:

The operation provides a means for the HLR to inform the SMS- C, which has previously initiated unsuccessful short message delivery attempts to a specific mobile station, that the mobile station is now recognized by the mobile network to be accessible. This service-center alert is accomplished via the use of the alert service-center mechanism.

4.3.3.2 SMS Message flow

SMS comprises two basic point-to-point services [10][11].

- Mobile Originated - Short Message (MO–SM)
- Mobile Terminated - Short Message (MT–SM)

4.3.4SMS Security

In the previous sections we described the Short Message Service characteristics. SMS is a useful technology for transmitting information to large numbers of recipients in a cost effective way. However, SMS is not secure enough for carrying financial transactions.

The same reasoning applies to SMS that applied to the USSD channel described in Section 5.2. There is no form of encryption or message integrity checking on the SMS message whilst traversing the GSM backbone. The situation is worsened by the fact that

SMS is a store and forward application and that SMS messages are stored on the SMS-C in clear. If an attacker gains access to the SMS-C, he can alter any SMS message. This means that an attacker can locate a SMS with a payment instruction and alter any part of the message with no one being the wiser.

Standard SMS technology can be used in other value-adding applications in the banking and financial arena to reduce associated risk. When used as a medium to provide real-time transaction history on a clients account, a financial institution can reduce its fraud risk. A bank might employ a SMS service to send SMS messages to its Credit Card holders each time a transaction is conducted on their Credit Card account. The cardholders will then instantaneously, or in a relative short period of time know that his Credit Card is being used for purchases not affected by him. This enables the cardholder to let the issuing bank know to cease all payment on his card, as he is not effecting the transactions. This reduces the attacker's window of opportunity for conducting fraud, and thus reduces the risk to the issuing bank as well as the cardholder.

4.4 THE SIM APPLICATION TOOLKIT (SAT)

4.4.1 Introduction

The SIM Application Toolkit (SAT), also referred to as SIM Toolkit (STK), is today mainly used as a tool that enables an operator-controlled menu for SMS and voice services. This allows operators to create specific applications resident on the subscribers SIM. It is also used for more advanced services that require high security, where the SIM plays a natural role as a Smart Card. The SIM Toolkit is, just as SMS, a well-proven GSM standard that has been out on the market since 1995. It has by now been incorporated into all major mobile telecommunications standards. Just like SMS it experienced a slow up-take in the beginning, partly as the market has been awaiting newer, more 'hyped' technologies.

The fact that the operator controls the SIM makes it an ideal platform for operator-provided services. The major drivers for implementing services using SIM Toolkit are the combination of its maturity, and its network technology independence as it is now incorporated in 3G and TDMA standards (GAIT). An application developed using SIM Toolkit will work in the 3G networks as well as when roaming into a foreign 2G network.

SIM Toolkit should be used for user-oriented services based on short transactions of the 'request – response' type and for implementing advanced SIM based functionality as a complement to a service developed u sing another browsing channel, e.g., the Web. SIM Toolkit is also ideal for server-initiated transactions, as the main data bearer is SMS. This makes it perfect for Internet services where the handset is only one of the devices, e.g., using the handset and the SIM for authentication of users and confirmation of payments while using a PC as the browsing device.

4.4.2SAT Operation

Figure 5.1 shows the entities involved in sending a SAT message from a sending application to a receiving application. Figure 5.1 also shows the entities where the security of the message should be applied.

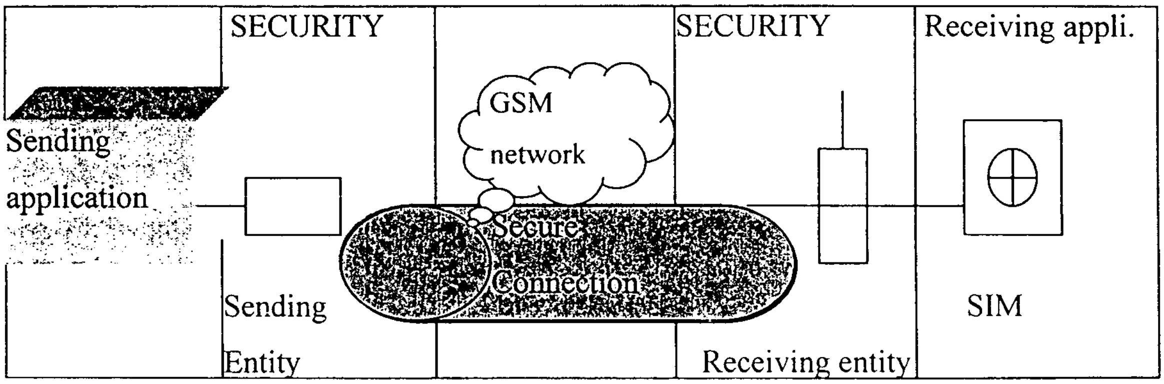


Figure 4.6: End-to-end System Overview of the SAT

The specification in [12] describes the flow of information between the sending and receiving entities and the security associated with that in the following manner and is depicted in Figure 4.7:

1. The Sending Application prepares an Application Message and forwards it to the

Sending Entity, with an indication of the security to be applied to the message.

2. The Sending Entity prepends a Security Header (the Command Header) to the Application Message. It then applies the requested security to part of the Command Header and all of the Application Message, including any padding octets.
3. The Receiving Entity subsequently forwards the Application Message to the Receiving Application indicating to the Receiving Application the security that was applied.
4. If so indicated in the Command Header, the Receiving Entity shall create a (Secured) Response Packet. The Response Packet consists of a Security Header (the Response Header) and optionally, application specific data supplied by the Receiving Application. The Response Packet will be returned to the Sending Entity, subject to constraints in the transport layer, (e.g. timing).

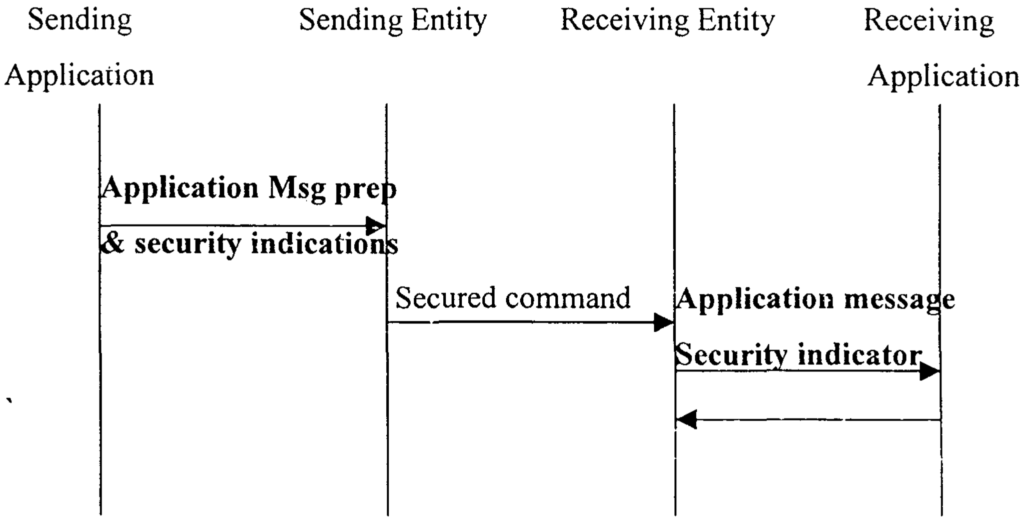


Figure 4.7: SAT Message flow

ELEMENT	LENGTH	COMMENT
Command packet identifier (CPI	1 octet	Identifies that this data block is the secured command packet
Command packet length (CPL)	Variable	This shall indicate the no.of Octets from and including the command header identifier
Command header identifier (CHI)	1 octet	Identifies the command header
Command header length (CHL)	variable	This shall indicate the no. of octets from and including the SPI to the end of RC/CC/DS
Security parameter identifier (CPI)	2 octets	See details
Ciphering key identifier (KI$_C$)	1 octet	Key and algorithm identifier for ciphering
Key identifier (KID)	1 octet	Key and algorithm identifier for RC/CC/DS
Toolkit application reference (TAR)	3octet	Coding is application dependent
Counter (CNTR)	1 octet	This indicates the no of padding octet at the end of secured data
Redundancy check (RC) Cryptographic Check (CC) Digital signature (DS)	Variable	Length depends on algorithm. A typical value is 8 octet if used and for a DS could be 48 or more.
Secured data	Variable	Contains the secured application message and possible padding octets

Table 4-1: Structure of the SAT Command Packet [5]

4.4.3 SAT Security

In [12], which is the international specification for SAT security, the authors state that there is a definite need to secure SAT related communication over the GSM network with a level of security chosen by the network operator or the application provider. No mandate is stated as to the level of encryption required, if any. Further to this, an assumption is made that the sending and receiving entities in the communication exchange are in secure environments.

This is basically equivalent to stating that there is no need to secure traffic traversing the Internet, for if we take the same point of view as the authors of [12], then when Bank A sends a file detailing client information and Credit Card details to Bank B over the Internet, we need not be concerned as both Bank A and Bank B are housed in secure environments.

According to [12], certain security requirements are to be satisfied when transferring Application Messages from the Sending Entity to the Receiving Entity. They are listed below with a brief description of why the service is needed and the mechanisms used to provide this service.

- **Authentication. :** Needed to authenticate the subscriber. This function is performed by the subscribers unique SIM.
- **Message integrity:** Needed to ensure that no corruption of the content of the message has occurred during transmission, be it accidental or intentional. This can be achieved by adding a Redundancy Check, Cryptographic Checksum, or Digital Signature to the security header.
- **Replay detection and sequence integrity:** Needed to protect the Receiving Entity against replay attack and Secured Packet Duplication. This is implemented by adding a counter in the calculation of the cryptographic checksum in the Security Header.
- **Proof of receipt and proof of execution:** Proves to the Sending Entity that the Receiving Entity has correctly received a secured packet, has performed the necessary security checks, and forwarded the contents of the secured packet to the receiving application. Proof of receipt is returned from the Receiving Entity in an acknowledgement to a secured packet transmitted by the Sending Entity. The acknowledgement takes the form of a Status Code in a response message.
- **Message confidentiality:** Needed to provide proof that the information contained in the messages exchanged is not disclosed to an unauthorized individual, entity, or process. This is achieved by encrypting the message with a block cipher.

As can clearly be seen from the previous sections, much development has gone into the security specification for the SAT. There are however some problems as the costs involved in implementing these solutions are high. This means that operators will have to implement some network changes in order to cater for these security requirements. Many of them just will just implement the basic changes needed to comply with the specification. As a result many operators cannot handle any message integrity checks on these SAT messages, which are pure SMS messages in essence, sent between the Sending and Receiving Entities.

For any financial institution this poses a serious problem. The same attack that was described in Section 4.2 and 4.3 still applies. Without a cryptographic message integrity check, there is no way to be sure that the message reaching the financial institution is actually the message their client intended to send. Even if some operators offer the encryption of the user's PIN for the m-Banking application, the attack is still possible. In order for us to explain this, let us once again look at an example:

Figure 5.8 shows a linear representation of the SAT Command Packet structure as described in Table 5.1. This is the makeup of the SMS send to the Receiving Entity during communication.

CPI	CPL	CHI	CHL	SPI	KI_C	KID	TAR	CNTR	PCNTR	RC\CC\DS	SECURE DATA

Figure 5.8: The SAT Command Packet Structure. [5]

The structure of the data section in a typical payment application is very easy to decode and would most probably look something like this:

0851234567112AB512012341234123450000857654321, where

- 0851234567 = the mobile number from which the transaction originated.
- 1 = Account Indicator.
- 12AB = encrypted PIN in Hex.
- 5120123412341234 = the account number the payment should go to.
- 50000 = the amount of the transaction.
- 0857654321 = the mobile number the confirmation message should be sent too.

Without message integrity checking, an attacker could easily alter the message to look Something like this (the field descriptions stay unchanged):

08512345671 12AB512011111111111150000 0852222222, with no one being any wiser.

In conclusion, it is very important to note that one cannot implicitly trust a network operator to implement the security required to utilize a channel for m-Commerce applications.

4.5 Wireless Internet Gateway (WIG) Technology

4.5.1 Introduction

The Wireless Internet Browser (WIB) technology was introduced to ease the development of SIM-based services using a generic SIM supplier independent interpreter on the card for any kind of application. The WIB optimizes the utilization of SIM memory in addition to offering a true interoperable execution environment on the SIM. It also solves the client/server problem since it uses standard web technologies. The WIB is contained within a 32k SIM card resident in a GSM phase 2+ compliant handsets.

The WIB operates together with a network component called Wireless Internet Gateway (WIG). The WIG opens up a channel to the WIB on the SIM. The WIG enables the usage of an easy to use application language, called the Wireless Markup Language (WML), for implementing SAT based mobile services. These messages are carried over the USSD or SMS channels. The channel of choice is however the SMS channel.

WIB is a specification for a SIM Toolkit interpreter, or browser, that is licensed free of charge to companies developing software for SIM cards. It has been in commercial use since the beginning of 1999 in various wireless applications, mostly in the m-Commerce area. In July 2001 there were approximately 25 million SIM cards on the market with the WIB enabled. Since mid 2000 there was a standardization initiative, the USAT Interpreter, within 3GPP to standardize the concept. With the WIB/WIG it is possible to implement ease-of-use services to the operator's installed base of mobile phones, and still be compliant with future technologies. With the WIB as a standard application in ROM from all SIM vendors all previous problems with SAT can be circumvented. SIM vendors have previously pushed for proprietary implementations with proprietary and difficult interfaces for service implementations. The WIB together with its corresponding delivery platform, WIG, drastically changes this as it provides one generic Internet based interface for service creation, independent of SIM suppliers.

4.5.2 WIG Architecture

The WIG Architecture in its simplest form is very basic in nature. A subscriber making use of the WIB sends messages via SMS to the WI G. The WIG converts these WML messages to a web based protocol like HTTP, or HTTPS if security is required, and forwards them to the content provider. The WIG Architecture is depicted in Figure 4. 9.

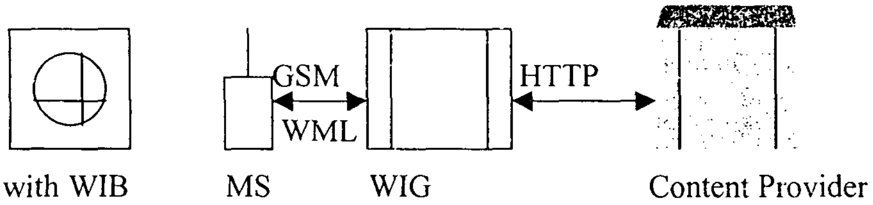


Figure 4.9: The WIG Architecture

4.5.3 Operation of WIG

WIG operates in conjunction with the SAT. The user browses a SAT menu and supplies the relevant information that is required to complete the transaction. Figure 4.10

illustrates a typical user interface and transaction flow on a mobile phone.

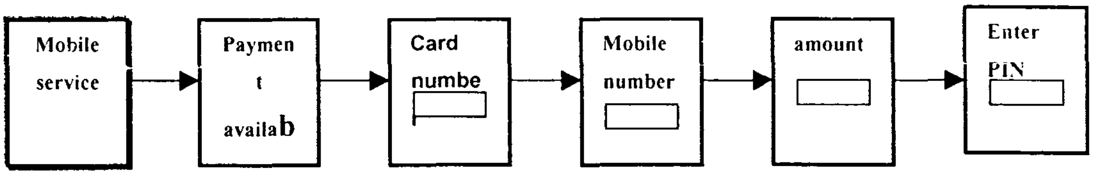


Figure 4.10: Example of a WIB payment menu on a GSM phone

Once the SAT application has all the information it requires to complete a transaction, it forwards the relevant information for processing to the WIG server. Figure 4.11 illustrates the transaction flow, followed by a brief description of each transaction.

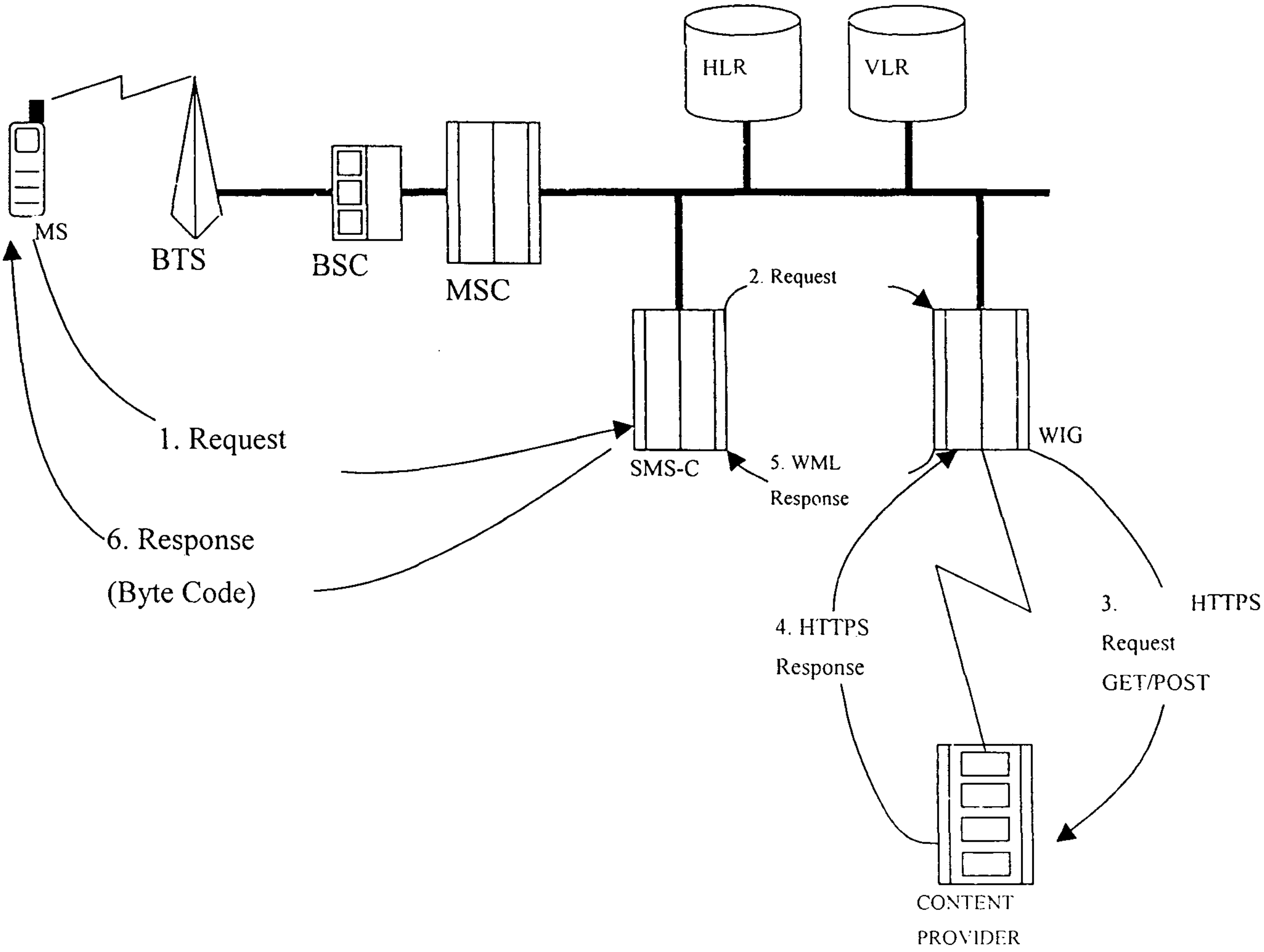


Figure 4.11: WIG Operation

The sequence of events during WIG operation is (Figure 4.11):

1. The Wireless Internet Browser (WIB) makes an URL request and sends it to the SMS-C via SMS.
2. The WIG Server receives the request from the SMS- C.
3. The WIG Server translates it into an HTTP GET or POST request and sends it to the Content Provider.
4. The Content Provider processes the request and sends an HTTP Response back to the WIG Server.
5. The WIG server passes the HTTP response and compresses the WML document into byte code and forwards the response to the SMS-C.
6. The SMS-C forwards the response to the handset. The WIB receives thc sequence of commands in byte code from the WIG server and runs these commands. The WIB will use SIM Application Toolkit for user interactive commands.

4.5.4 Security of WIG

WIG in itself does not offer any additional security features apart from those offered by the technologies discussed in preceding sections. WIG is merely a standard that allows any handset with a SAT application loaded, to interact with a content provider.

Chapter 5

PROPOSED MODEL FOR M-COMMERCE BASED ON WIG

5.1 Introduction

In the previous chapters various technologies that can provide the tools to power m-Commerce applications were discussed. In this section these elements will be combined to present a model that will provide a secure m-Commerce application that relies on WIG technology.

It is of cardinal importance for any financial institution wishing to utilize new technologies to afford their customers with an enhanced capability, to ensure some basic security concepts. Some of these concepts include:

1. **Authentication of the client and the Financial Institution**. Making sure that the person initiating the transaction is actually authorized to do so, and that he is communicating with the institution he intends to.
2. **Confidentiality of client information**. Ensuring that a client's Credit Card details or home address does not get exposed to unauthorized persons, organizations or applications.
3. **Integrity of payment instructions.** Ensuring that the payment instruction received from a client is actually the instruction he wishes to have processed, and not an instruction altered during transmission.
4. **Non-repudiation.** Ensuring that the client cannot deny ever sending the instruction to the institution.

In this section we will look at the practical implementations of transmission security in the financial sector, with specific reference to the way organizations like VISA and MasterCard maintain integrity and confidentiality of transmissions. A model is then proposed that caters for these requirements, either by means of cryptography, or by some other means like business processes. The author makes the assumption that no entity in the end-to-end scope of the solution, except the financial institution itself, can be trusted to implement security on their behalf. This does not imply that the institution should re-invent the wheel, but rather that it should understand what the technology can guarantee and cannot be guaranteed. Where the technology cannot provide an acceptable solution, a creative alternative should be sought.

5.2 Message Confidentiality In The Financial Sector

The following section defines the process for ensuring message integrity and confidentiality used by all major banks and Credit Card providers throughout the world.

Financial institutions make use of symmetric key encryption to secure Personal Identification Numbers (PIN), or any other information deemed sensitive enough, traversing the information networks. The scheme they employ enables an Acquiring bank to communicate with any Issuing bank in the world, making use of well defined processes for symmetric key exchanges between banks and 3rd parties. This scheme is illustrated in Figure 5.1.

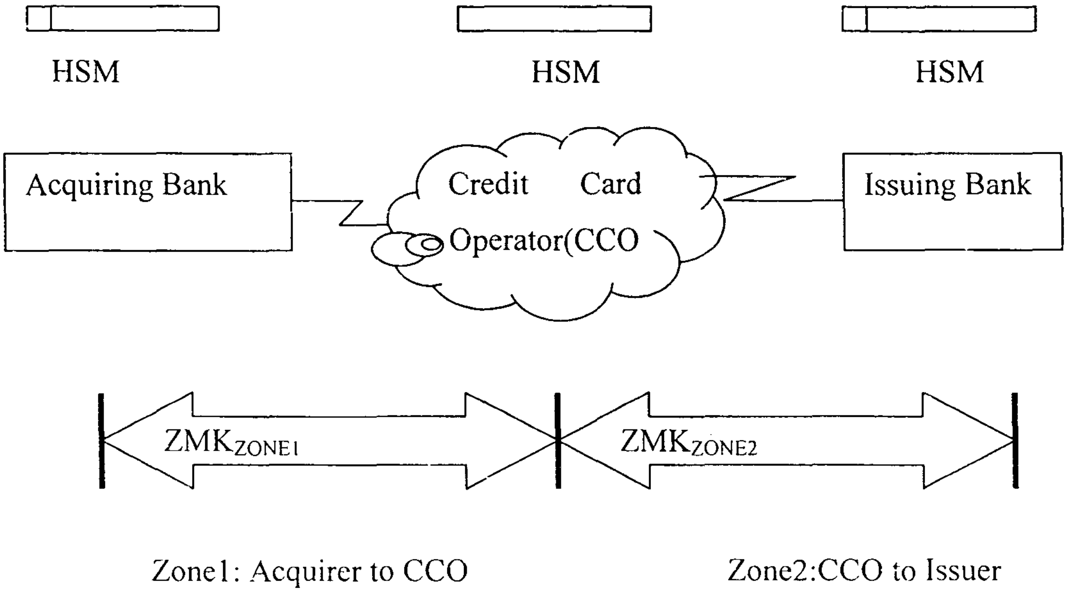


Figure 5.1 Key Exchange Architecture

The Acquiring and issuing institutions exchange a Triple DES Zone Master Key (ZMK) or Key Encrypting Key (KEK), also called a Zone Control Master Key (ZCMK), with the Credit Card Operator (CCO), like VISA or MasterCard. This key exchange is done via a paper-based method, with three different key components of the key being couriered to three different individuals within the CCO. This ZMK is then stored in an encrypted format via a machine specifically designed for banking encryption. This machine is called a Host Security Module, or a Hardware Security Module (HSM). These machines comply with the strictest security standards and are Federal Information Processing Standard (FIPS) 60-1 Level 3 certified. The ZMK is used to encrypt all subsequent key exchanges between organizations.

As an example, these organizations will also exchange a key called the Zone PIN Key (ZPK), which is used to encrypt all PINs for transmission between organizations. Either the financial institution or the CCO generates the ZPK. It is then encrypted under the 3DES ZMK and sent electronically to the other party. It is important to note that the financial institutions do not exchange keys with each other, but merely with the CCO. The CCO keeps a database of all the banks that makes use of their services, and all the keys that the bank hold s. These entire keys are stored encrypted under that specific bank's ZMK. They are not stored in clear text. In order for the bank and the Credit Card institutions to communicate with each other, all communicating parties have to make use of the same encryption standards and technologies. For this reason all the banks also by default posses HSM processors.

The HSM makes use of hardware encryption to translate a message from being encrypted under one key (the Acquirer), to being encrypted under another key (the Issuer). These translations take place in the hardware and under no circumstances is any part of the decrypted message visible to anyone apart from the hardware. The actual functioning of these devices falls outside of the scope of this article.

Figure 6.2 depicts this translation process graphically. (VISA is only used as an example)

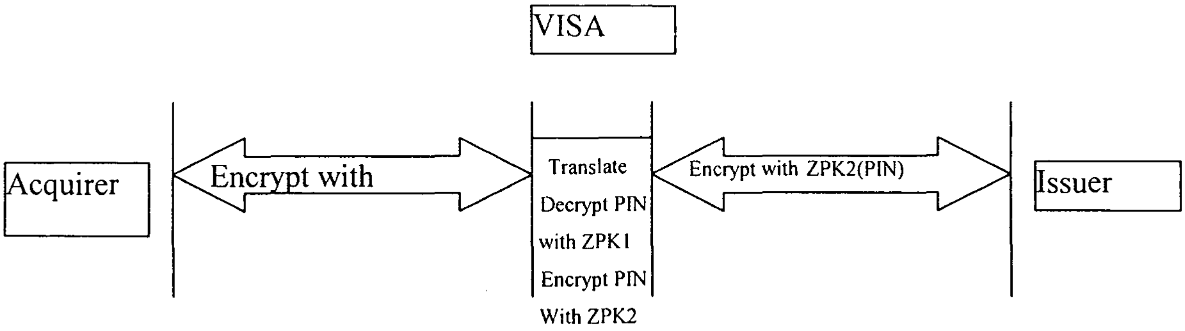


Figure 5.2: PIN Translation

However, there is one problem that has been identified. If only a PIN is encrypted, problems could arise with the use of cryptographic padding of the PIN before encryption, as PIN lengths differ. To overcome this problem, banks make use of either one of two modes of operation, defined in the international standard for PIN block generation. The ISO 9564-1 standard defines modes of PIN block generation. The most widely used is either the Format 0 PIN Block, or the Format 1 PIN Block. For the purpose of this dissertation, and the proposed solution, we will make use of the Format 1 PIN Block, although Format 0 could just as easily be used.

As an example, the ISO 9564-1 Format 1 PIN block is generated in the following manner:

For the I SO 9564-1 Format 1 PIN block, the PIN is formatted as follows:
61234RRRRRRRRRR, where:

- 1 = ISO Format indicator
- 4 = PIN Length
- 1234 = PIN
- RRRRRRRRRR = Padding with random Hex values

For 5 digit PINs it will look like: 612345RRRRRRRRR and for 6 dig it PINs it will look like: 6123456RRRRRRRR

For transportation, the PIN Block now needs to be encrypted with the ZPK shared between the Acquirer and the CCO.

The CCO then translates the received encrypted PIN Block to being encrypted under the ZPK shared with the Issuer of the card.

By making use of this process, the banking fraternity has solved the problem of encryption of PINs between Acquiring and issuing banks. As this process has been proven to function admirably in the real world, the authors chose to apply this architecture to the proposed m-Commerce model.

5.3 Client Registration

Client registration plays an integral part in the solution. The more information that can be obtained from the client during registration, the less information about the client needs to be transmitted across the network. The client can register via a secured website, where after the information should be confirmed with the client via some other channel, (e.g. e-mail, telephone call, etc.) Assuming the client is already an account holder with the institution, additional information that should be collected includes:

- The client' s mobile number or any other mobile number he wishes to have access to this service.
- The account number(s) of the account(s) he wishes to use during his transactions.
- The payment limits he wishes to implement. This could be in the form a specified maximum amount to any other account, any amount payable to only certain account(s), or a combination of both. These limits might also depend on the business rules imposed by the financial institution. If they decide to only allow payments to pre-defined accounts whilst using this service, then the service is sold as such.
- Any other information the institution might feel is pertinent. Some institutions might want the client to sign an agreement that specifies that the client carry the risk associated with using the service.

Once the client is registered, his information should be stored in a database that can easily be accessed by the content server. In this database, the client's mobile number should be linked to his account number, so that either can easily identify him. If the client wishes to use more that one account for these payments, each account should be assigned an identifier. This identifier can take any form, i.e. an integer number, or an alias. This will ensure that the client's account details are never transmitted across the network.

5.4 Confidentiality Of The Customer PIN

Each 32K SIM card distributed by the Mobile Networks has a set of unique double-length DES keys. These keys are diversified and unique per device. In order for the financial institution to verify a PIN encrypted with one of these keys, they either need access to the key used for encrypting the PIN, or the PIN needs to be translated to a key that the financial institution knows. The network operators limit usage of these keys and they are very reluctant to give one access to these keys. Despite this fact, they have a system in place for encrypting certain sensitive information from the MS to the HSM on the backend using one of these keys.

The end-to- end confidentiality of the PIN is ensured by a combination of standard financial processing and encryption by the network operator. This process is explained in Figure 3.

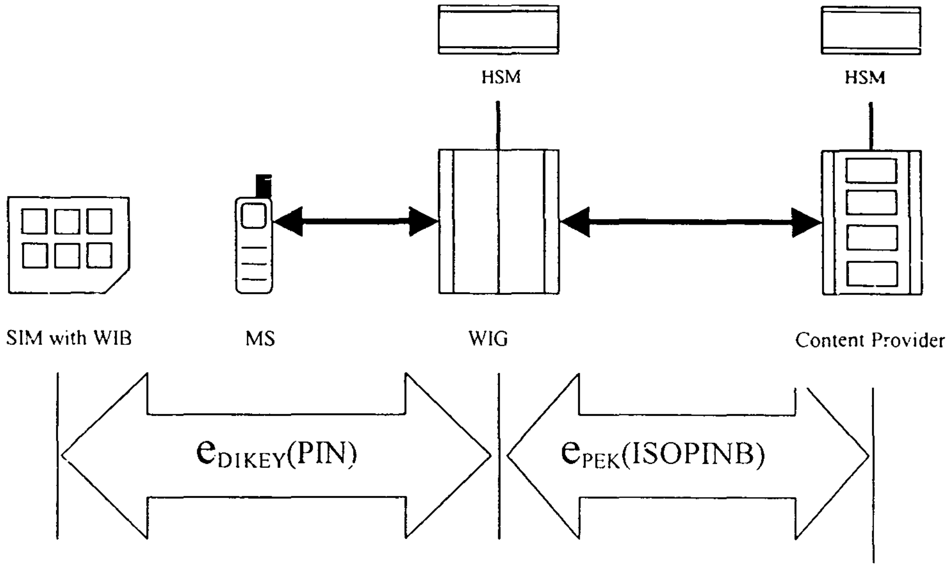


Figure 5.3: Confidentiality of the PIN

Once the client has completed the entering his PIN, the PIN is encrypted using 3DES with one of the unique keys on the SIM and sent, along with the rest of the message, to the WIG server. Once the PIN arrives at the WIG, the HSM is called to decrypt the PIN, reformat it into an ISO 9564-1 Format 0 PIN Block and re-encrypt it under a 3DES Pin Encryption Key (PEK). All the cryptographic functions should be performed in a tamper evident hardware security module (HSM). This PEK is a key that is securely exchanged between the network operator and the financial institution.

The reformatted encrypted PIN block is then sent through to the financial institution where the PIN is verified against an encrypted PIN stored on the financial institution backend, using a one-to-one compare method. This function should also be performed within a HSM. This process should ensure the end-to-end security on the clients PIN.

5.5 Authorization Of The Financial Institution

The question arises: "How does the client know that he is actually communicating with his own bank?" The answer is quite simple. By making use of a Personal Assurance Message (PAM) the client can have full comfort that he is actually communicating with his desired bank. The PAM is a phrase recorded by the financial institution upon client registration. It is a shared secret that only the bank and the client know. It can take any form, like "THESIS IS COMPLETE", or any other phrase that falls within the criteria specified by the financial institution. This PAM is presented on the client's mobile phone once the communication keys are passed back to him. This message which only the bank and the client knows, then assures the client that he is in fact communicating with his bank.

5.6 End-To-End Message Integrity

There are two distinct possibilities in solving the problem of message integrity in mobile payments solutions by using the current technology available to us. The first option makes use of the same architecture as used in securing the confidentiality of the client's PIN. In this scenario, one of the derived keys resident on the SIM card is used to do a 3DES CBC MAC of the whole message, excluding the encrypted PIN. This message is then packed into the SMS and sent to the WIG. Once the SMS arrives at the WIG, the application strips the PIN from the SMS and verifies the MAC on the remainder of the message. If the MAC verified correctly, the WIG application translates the PIN into the ISO format PIN Block. The WIG application then recalculates the MAC on the entire message, including the encrypted PIN Block, using the Message Authentication Key (MAK) that was securely exchanged with the financial institution. This new message is then sent to the financial institution, which verifies the MAC and the client PIN before acting on the payment instruction contained within the SMS. This architecture is depicted in Figure 4.

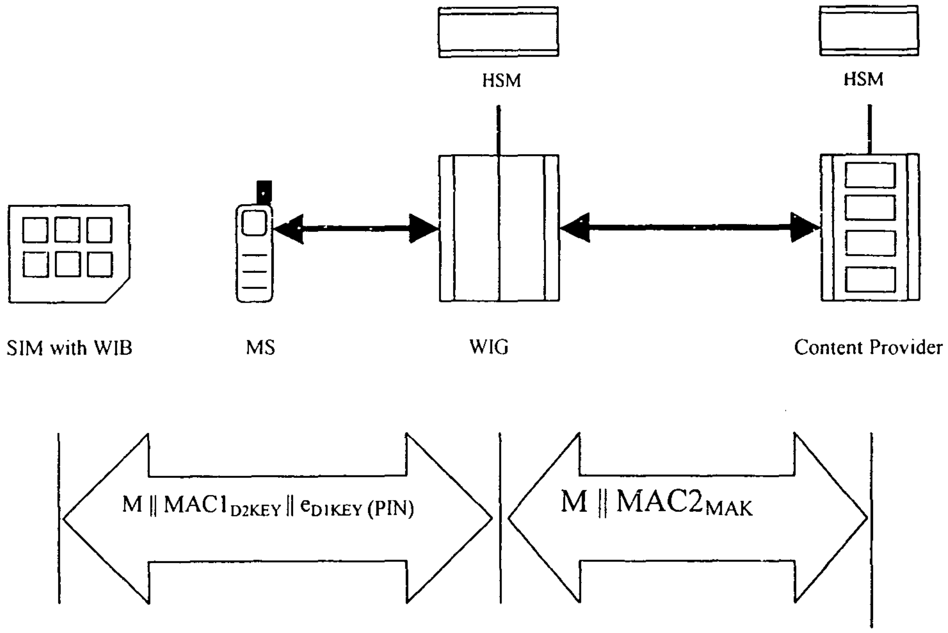


Figure 5.4: Message Integrity with MAC translate

Although this solution is possibly the easiest, it is not always possible to implement it. Some network operators have proprietary HSM modules, which cannot do this kind of translation without development work. As this is highly specialized equipment, any development work required is very costly. Network operators are reluctant to incur such costs. Some network operators do not even feel that the need to validate the integrity of the message from end-to-end is justified, as they are of the opinion that their networks are secure. The second option entails the use of a unique message authentication key per transaction.

In this model, the financial institution and the network operator establish a secure Zone Master Key, also called a Key Encryption Key, between the two HSM modules. The application resident on the client's handset now has to cater for an online "registration" process before the actual payment application can take place.

The client application sends a SMS to the WIG server indicating that the client is ready to perform a payment transaction. The WIG then send the request to the financial institution' s application server which in turns requests the financial institution's HSM to generate a random session message authentication key (SMAK) encrypted under the secure ZMK exchanged previously. The financial application server forwards the encrypted key to the WIG who asks its HSM to translate the encrypted SMAK to a key resident on the subscribers SIM card. The translated key is then sent to the SIM, where it is used to generate the MAC on the message, without the encrypted PI N, before it is packaged into the SMS.

Upon receiving the SMS, the WIG server translates the encrypted PIN into the ISO format PIN Block before sending the payment instruction to the financial institution. The network operator does not have to translate the MAC to a different key, as the financial institution has the SMAK, and can therefore verify the MAC on the payment instruction by merely removing the encrypted PIN block from the message. This provides true end-to-end message integrity verification. Figure 5.5 depicts this.

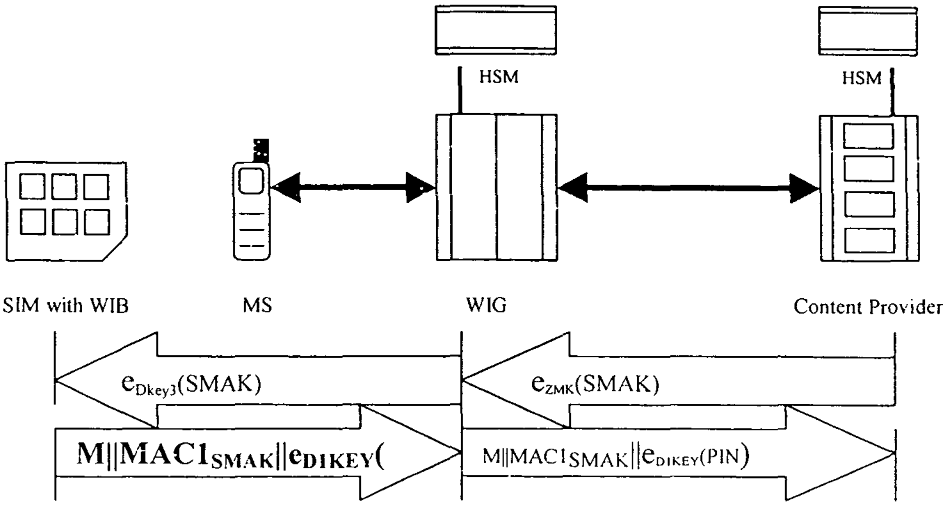


Figure 5.5: End-to-End Message Integrity

5.7 The Architecture Of The Proposed Solution

The proposed solution will have a very similar architecture to WIG, as this is the technology used. Figure 5.6 illustrates this. The only difference here is that the Content Provider is indicated as a cloud.

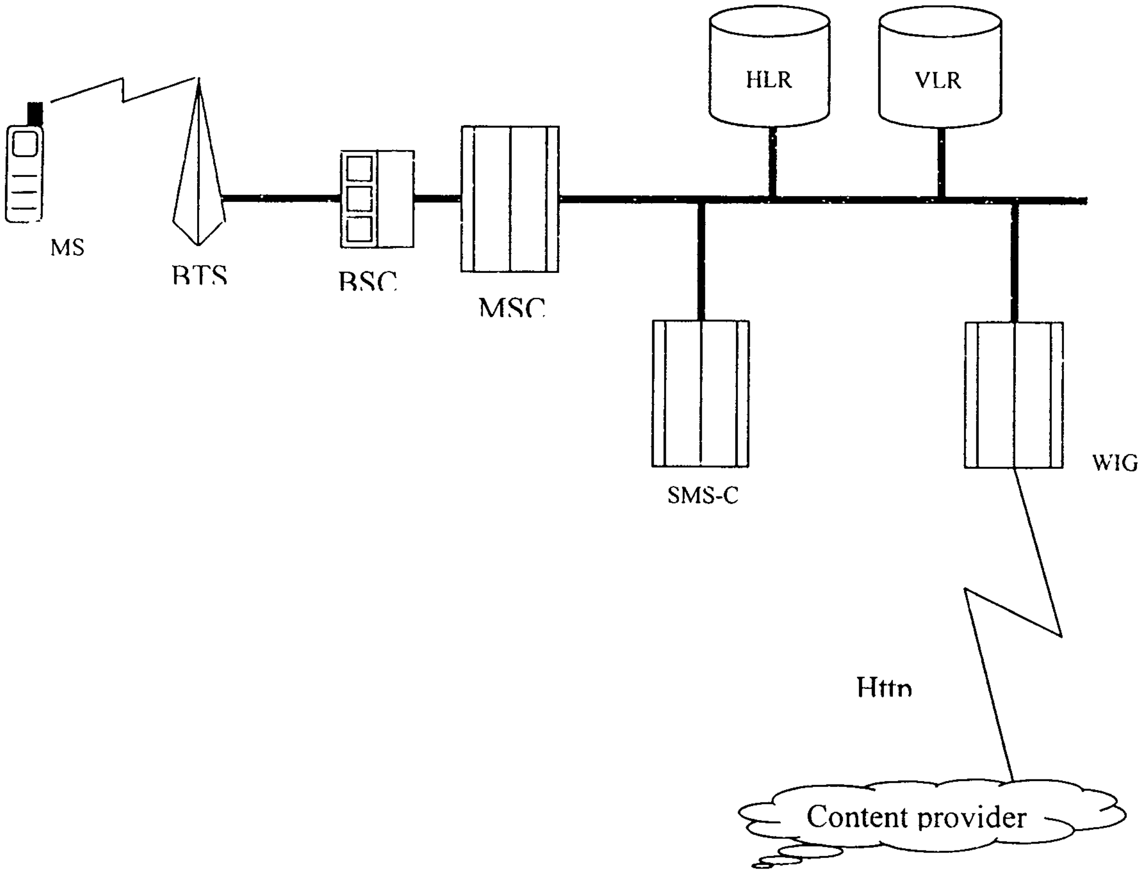
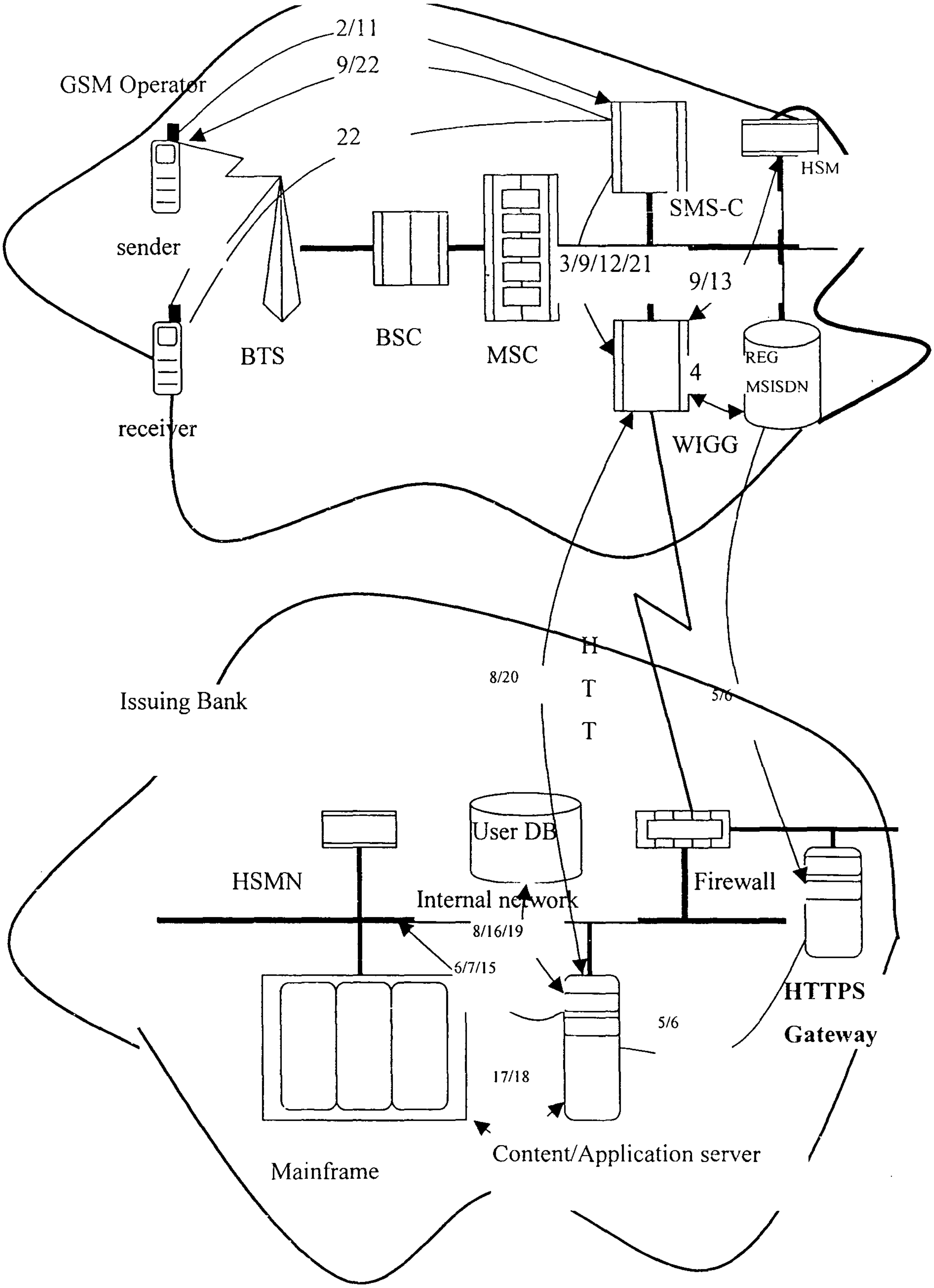


Figure 5.6: Architecture of the Proposed Solution



5.8Transactional Flow

Figure 5.7 shows the flow. Each step is discussed in sequential order.

1. The flow of a typical transaction in the proposed model would be:
2. Client selects the WIG banking application on the phone menu and selects the initialize application.
3. The message is sent via SMS to the SMS-C.
4. The SMS-C forwards the message to the WIG server.
5. On receiving the message the WIG gateway verifies the MSISDN is registered for the application.
6. If the client is registered, the WI G server sends the MSISDN of the user to the Application Server at the bank via the HTTPS gateway, and requests a random Message Authentication Key (MAK) from the Application Server, which will be used to compute a MAC on the transaction message.
7. The Application Server makes a call to its HSM requesting a MAK.
8. The HSM responds to the Application Server with the MAK encrypted under the ZMK. The Application Server retrieves the user's PAM and forwards it and the encrypted MAK to the WIG server via HTTPS.
9. The WIG server decrypts the MAK and sends it and the PAM to the handset via an instruction message.
10. Upon receipt of the instruction message, the SAT application launches the WIB and displays the PAM to the client. The client is prompted to verify the PAM. Upon Successful verification of the PAM; the SAT prompts the user to enter the required information, i.e. from account indicator, to account number, amount, and the mobile number where the confirmation message should be sent. After displaying a confirmation message, the user is required to enter his PIN. The PIN gets 3DES encrypted with the unique derived keys resident on the SIM card. The SIM constructs the message to be sent, appends a timestamp and performs a DES MAC on the whole message, excluding the encrypted PIN and the timestamp.
11. The handset forwards the message to the SMS-C
12. The SMS- C forwards the message to the WIG server.
13. The WIG server, making use its the HSM, translate the encrypted PIN to an ISO

9564-1 Format 0 PIN block.

14. The WIG appends the PIN block and the users MSISDN to the original message, excluding the 3DES encrypted PIN, and MACs the message with the ZAK shared with the bank. The WIG server sends the message to the Application Server via the SSL link.
15. Upon receipt of the message, the Application Server verifies the MAC on the message from the WIG. It also verifies the client's PIN using the PIN block, and verifies the MAC of the original SMS.
16. If all verifications succeed, the Application Server retrieves the client's account number from the user database, and formulates the payment instruction to be sent to the mainframe.
17. The Application Server sends the payment instruction to the mainframe.
18. The mainframe processes the transaction and sends a response code to the application server.
19. The Application Server MACs the response code by using the ZAK.
20. The application server sends the MAC as a confirmation code, along with details of the transaction to the WIG indication the MSISDN numbers that should receive the confirmation message.
21. The WIG for wards the confirmation message to the SMS- C.
22. The SMS-C forwards the confirmation message, as a normal SMS to the intended recipients.

In this proposed model, data integrity is kept by means of a MAC from the handset to the issuing bank. This kind of end-to-end integrity checking is required in order to secure this channel for m-Commerce applications. The confidentiality of data in this instance is not so crucial, as very little information about the client is transmitted across the network. This is why client registration plays a very important role in the proposal.

Chapter 6

ANALYSIS OF THE PROPOSED SOLUTION

6.1Summary of the performance of proposed solution

ATTACK	Before application	After application
Interruption	Y	Y
Interception	Y	Y
Modification	Y	N
Fabrication	Y	N
Confidentiality	Y	Y
Message integrity	Y	N
Authentication	Y	N
Replay	Y	N
Non-repudiation	Y	N

WHERE
Y: VULNERABLE TO ATTACK
N: Not Vulnerable To Attack

6.2 Possible Solution Vulnerability

6.2.1 The WAP Gap

Although Wireless Transport Layer Security (WTLS) provides us with Wireless Application Protocol (WAP) security over the wireless network, much the same as SSL does in the wired medium; a huge flaw exists in some implementations of WAP. This is commonly referred to as the WAP gap . The WAP gap comes from the manner in which the WAP Gateway is implemented. Figure 6.1 shows the WAP gap.

The WAP gap highlights the issue of control over the WAP Gateway. In certain implementations of the WAP model the WAP Gateway is not under the control of the financial institution, etc. In essence, not having control over the physical and logical security of the gateway renders it a un- trusted element in the end-to-end security of the transaction data.

Due to the WAP stack functionality it is not feasible to do away with the gateway. It is therefore necessary to establish the extent of the risk and alternatives to of address these risks.

Some might argue that the WAP gateway is not a security risk because the gateway vendors are aware of the issue and have taken steps to ensure that the process of decrypting from WTLS and re-encrypting into TLS cannot easily be compromised. Typical steps taken will be to ensure that the decryption and re-encryption takes place in memory, that keys and unencrypted data are never saved to disk, and that all memory used as part of the encryption and decryption process is cleared before being handed back to the operating system.

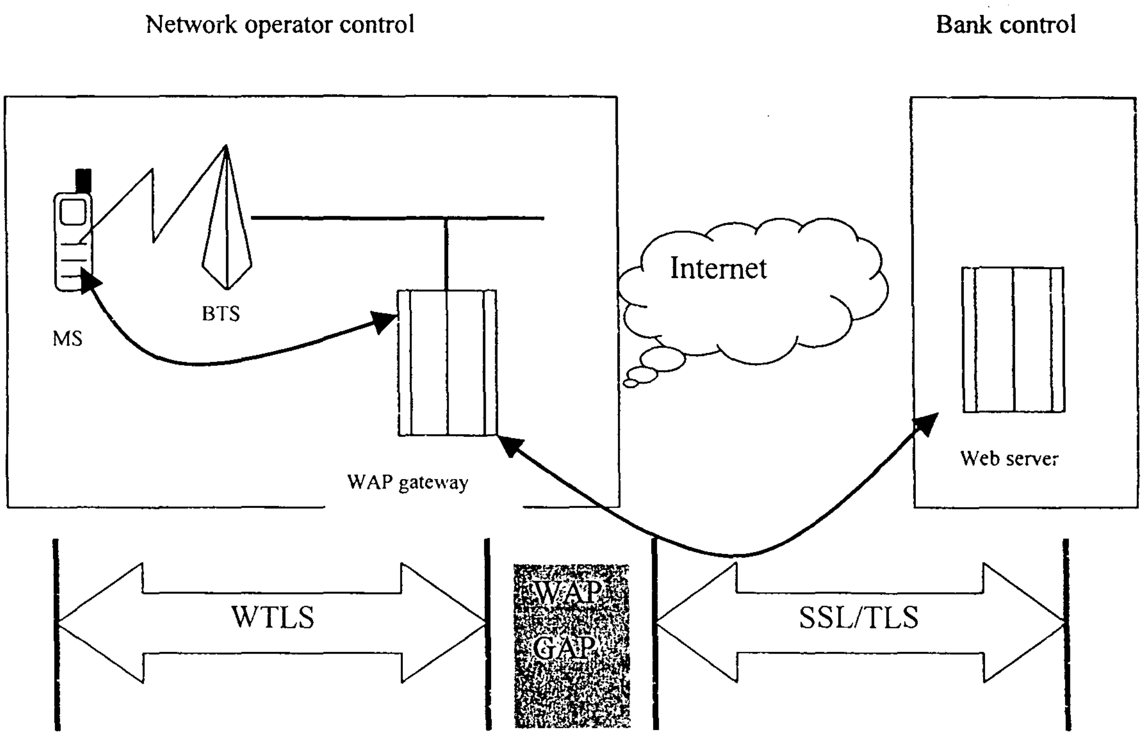


Figure 6.1: The WAP Gap

The problem with this is that there are no standards or guarantees about these precautions. There exists no way of ascertaining how robust a vendor's implementation actually is, and in the case of a gateway that is hosted by a network operator you may not even be able to tell who implemented it. The possibility exists that a vendor that can implement a WAP gateway on a very secure operating system in a thoroughly secured environment under the control of extremely competent administrators could provide a reasonably secure implementation. Even so, there is still an exposure around the gateway and at some stage, when it can become financially feasible, it will become a target for hacker.

With a minor change in the network architecture of Figure 5.7, we can secure WAP to such an extent that the risk becomes either manageable or acceptable to the organization. By moving the gateway "in-house" to the financial institution, the WAP gap becomes manageable. Figure 6.2 shows this architecture.

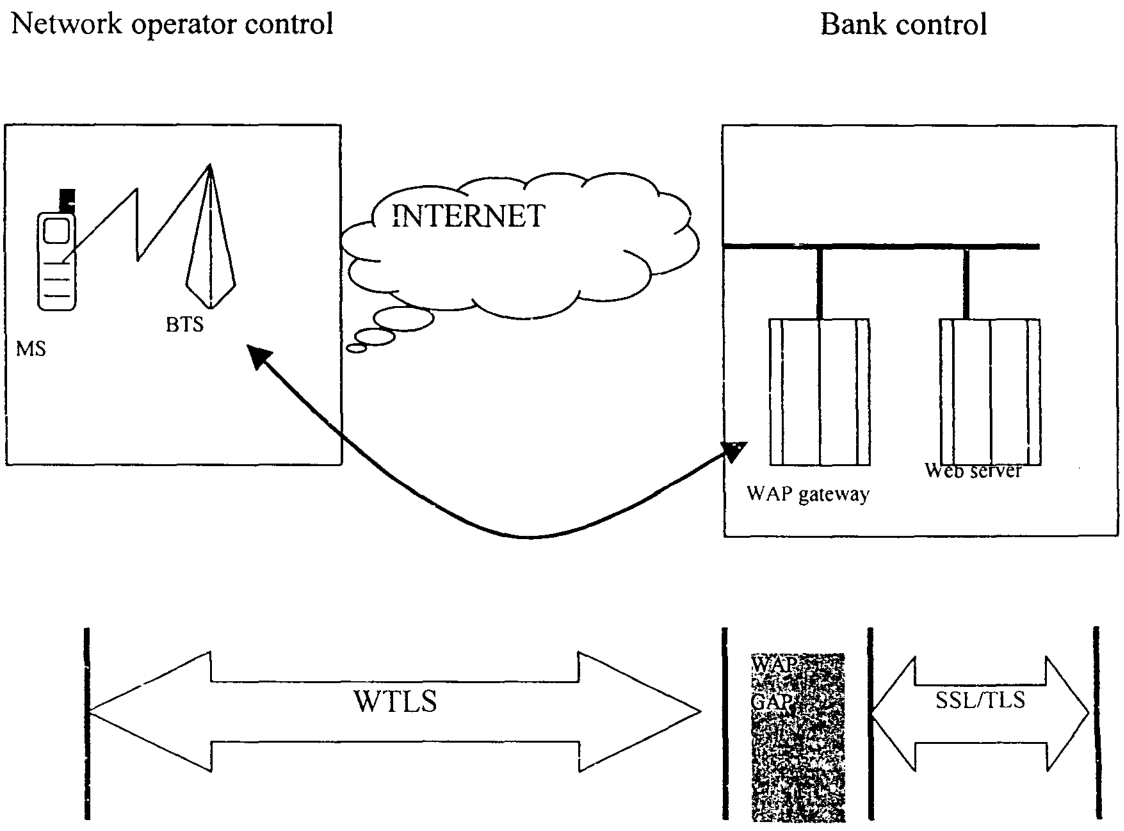


Figure 6.2: Closing the WAP Gap

6.2.2The WAP GAP applied to WIG

In analysis of the proposed solution, the question arises: “Does the WAP Gap principal apply to the proposal?” In a WTLS secured WAP application, the WAP server has to convert the WTLS instructions received from the handset to SSL instructions in order for the Web server to understand them. This implies that the WAP server has to “translate” the WTLS encryption into SSL encryption. It would therefore seem that the same vulnerability could exist in the proposed WIG solution due to the fact that the WIG server has to translate the MAC from encryption under one key to the next. Does this not then constitute a WIG Gap? Figure 6.3 show this graphically.

There is a difference that needs to be noted. In the WAP scenario, the “translation” takes place in the WAP server’s memory, and the local resources of the WAP server is used to compute the translations in encryption schemes. In the WIG scenario, the translation takes place in an industry certified hardware encryption device as discussed in section 6.

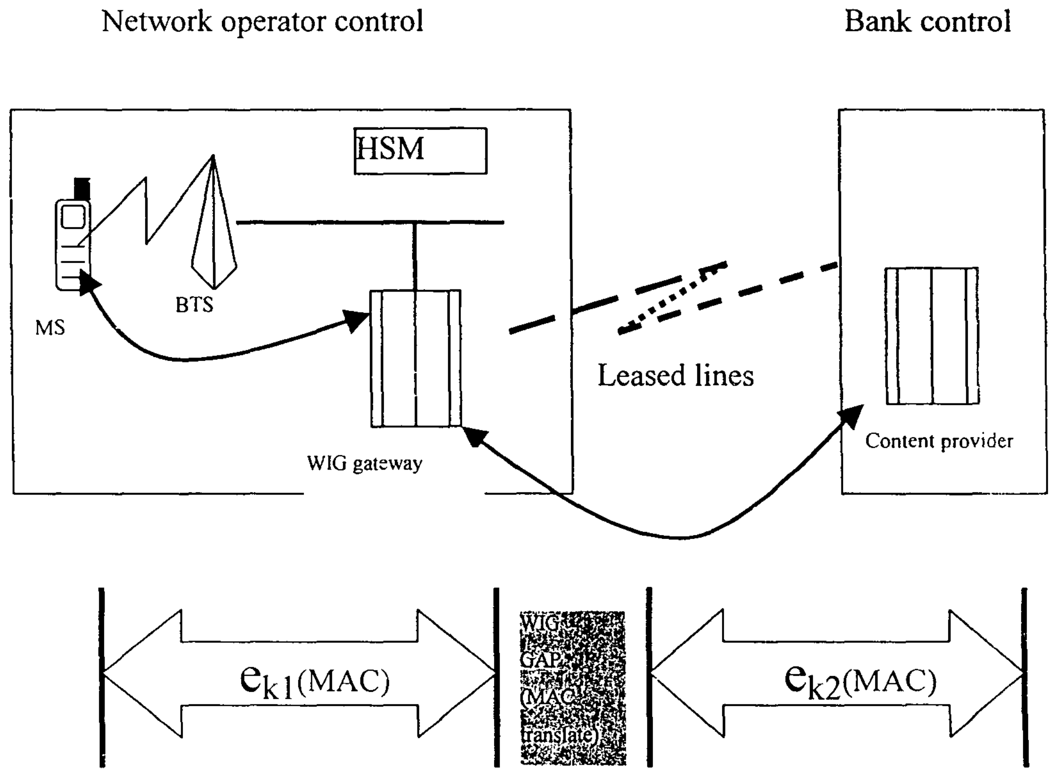


Figure 6.3: The WIG Gap

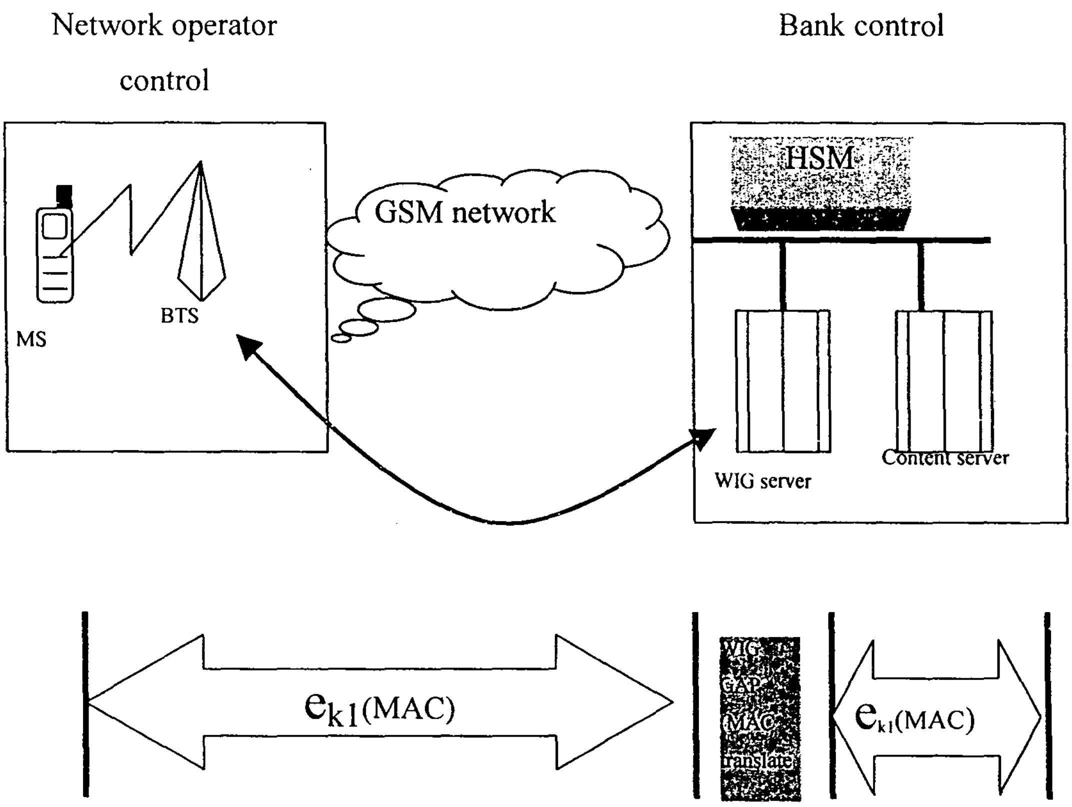


Figure 6.4: Closing the WIG Gap

These hardware encryption devices are FIPS 60-1 level 3 certified, which by implications means they comply with a certain standard. This standard is accepted by all financial institutions in the world, including the four biggest Credit Card Organizations, i.e. VISA, MasterCard, American Express and Diners Club. It can therefore be assumed that, given the architecture of the proposed solution, the use of a HSM in the design greatly reduces, or even negates, the WIG Gap vulnerability as described above. The only problem the author has with this scenario is that the HSM is not under control of a trusted entity. It is one thing to trust that a Credit Card Organization has implemented their HSMs correctly, and another matter entirely to trust a Network operator to implement the HSMs in the specified manner. The author suggests that this can be achieved by certifying the Network operator for m-Commerce applications. In doing so, the implementation of the WIG server and its corresponding HSM can be audited by independent individuals in order to insure that the devices are configured as specified.

Another option would be to move the WIG server into the trusted entity's area of influence and control. This would imply that as in the WAP Gap, the WIG server gets moved into the Network of the bank, as displayed in Figure 6.4.

This change in architecture effectively negates the WIG Gap vulnerability. This option is extremely expensive, due to the nature of the WIG server, and the volumes such a server should be able handle. Another factor that adds to the cost is that a bank wishing to deploy this solution would require at least one WIG server per Network operator. In India this would imply that each bank that wants such a mobile payments solution would require many WIG servers in order to cater for the clients of the many Network operators resident in India.

In the opinion of the author, the added security benefit derived from deploying the WIG server internal to the bank's network does not warrant such a huge expense.

Chapter 7

CONCLUSION

GSM was originally designed to offer voice services, however today many of the value added features have been added. Some of these features are Short message service (SMS), Wireless Internet gateway (WIG) SIM application Toolkit (SAT). Each of these application brings with them host of opportunities and applications. M-commerce is one such application. By using mobile phones over a GSM network users of these applications can effect a financial transactions from any where in the world where there is GSM network

In order to ensure the integrity of these messages must be secured .GSM security was originally designed for voice, which cannot be trusted for financial payment because of huge risk involved.

Due to the cost factor lot of operators will not be ready to take the burden hence the simple SMS message with SIM applications toolkit is analyzed wherein the author proposed that the banking pin was encrypted by making use of the SIM card resident on the mobile station which ensures confidentiality of the PIN. All other data except the PIN was then fed through a 3DES algorithm and a MAC was derived from it. This MAC accompanied the SMS message from the mobile station to the financial institution providing end-to-end message integrity. As a result financial transactions can be conducted over Wireless Internet Gateway with confidence.

7.1 Future work:

The advent of newer technologies like GPRS, 64 K SIM Card , 3rd generation mobile phones and SMART phones promises much more than what we have right now. Newer Technologies will open the Pandora's box in terms of newer challenges, which I am looking forward to.

BIBLIOGRAPHY

[1] Senn, J. "The Emergence of M-Commerce", IEEE Computer Magazine. December 2001,pages 148-150.

[2] Varshney, U., Vetter. R, Kalakota, R. "Mobile Commerce", A new frontier, IEEE Computer Magazine.

[3] Duraiappan. C, Zheng. Y. "Enhancing Security In GSM". University of Wollongong, Melbourne Australia.1999.

[4] Schmidt, M. "Consistent M-Commerce Security on top of GSM -Based Data Protocol-A Security Analysis" University of Siegan, Institute for data communication system. Siegen Germany.2001.

[5] European Telecommunications Standards Institute. Digital Cellular telecommunications system (phase 2+), "Security mechanism for the SIM application toolkit" stage 2.GSM 03.48 version 6.0.0 Release 97, ETSI, April 1998.

[6] Stallings, W. "Cryptography and Network Security, Principles and Practice". Second Edition. Prentice Hall, 1999.

[7] Margrave, D. "GSM Security and Encryption", George Mason University.

[8] Scourias, J. "Overview of GSM: The Global System for Mobile Communications". University of Waterloo March 1996.

[9] Sema, "Unstructured Supplementary Services Data (USSD)- A detailed overview".

[10] The International Engineering Consortium. "Wireless short message service (SMS)" Web ProForm Tutorials.

[11] S. Feldman, "Mobile Commerce for the masses", IEEE Internet computing 4(6)(2000)(June)) 74-75.

[12] Buckingham, S. "Success 4 SMS White paper", Mobile Streams 2001

[5] European Telecommunication Standards Institute. "Point to Point Short Message Service Support on Mobile Radio Interface", GSM 04.11 ETSI January 2003.

[14] European Telecommunication Standard Institute: Digital Cellular Telecommunication System (phase 2), "Technical realization of Short Message Service (SMS) point to point (pp)" GSM 03.14 ETSI October 19996.

[15] Goldberg, Briceno & Wagner: "An implementation of the GSM A3A8 algorithm".

[16] Goldberg, Briceno& Wagner: "A pedagogical implementation of the GSM A5/1 and A5/2`voice privacy` encryption algorithm".

[17] Biryukov, Shamir & Wagner: "Real Time Cryptanalysis of A5/1 on a PC".