

**NCS TIB 03-2**



---

---

**TECHNICAL INFORMATION BULLETIN 03-2**

---

---

**SMS over SS7**

**NATIONAL COMMUNICATIONS SYSTEM**

**December 2003**

OFFICE OF THE MANAGER  
NATIONAL COMMUNICATIONS SYSTEM  
701 SOUTH COURTHOUSE ROAD  
ARLINGTON, VIRGINIA 22204-2198



# SMS Over SS7



**Office of the Manager  
National Communications System**

**December 2003**

**Communication Technologies, Inc.  
14151 Newbrook Drive, Suite 400  
Chantilly, Virginia 20151  
703-961-9088 (Voice)  
703-961-1330 (Fax)  
[www.comtechnologies.com](http://www.comtechnologies.com)**

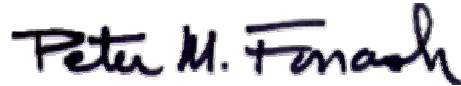


SHORT MESSAGE SERVICE OVER SIGNALING SYSTEM 7

December 2003

PROJECT OFFICER  
PUBLICATION:

APPROVED FOR



DALE BARR, JR.  
Sr. Electronics Engineer  
Technology and Programs Division

PETER M. FONASH  
Chief, Technology  
and Programs Division

FOREWORD

Among the responsibilities assigned to the National Communications System, is the management of the Federal Telecommunications Standards Program. Under this program, the NCS, with the assistance of the Federal Telecommunications Standards Committee identifies, develops, and coordinates proposed Federal Standards which either contribute to the interoperability of functionally similar Federal telecommunications systems or to the achievement of a compatible and efficient interface between computer and telecommunications systems. In developing and coordinating these standards, a considerable amount of effort is expended in initiating and pursuing joint standards development efforts with appropriate technical committees of the International Organization for Standardization, the International Telecommunication Union-Telecommunications Standardization Sector, and the American National Standards Institute. This Technical Information Bulletin presents an overview of an effort which is contributing to the development of compatible Federal and national standards in the area of national security and emergency preparedness (NS/EP). It has been prepared to inform interested Federal and industry activities. Any comments, inputs or statements of requirements which could assist in the advancement of this work are welcome and should be addressed to:

National Communications System  
Attn: N2  
701 S. Court House Road  
Arlington, VA 22204-2198



# Table of Contents

|  |      |
|--|------|
| Executive Summary .....  | ES-1 |
| 1. Introduction.....   | 1    |
| 2. Background.....   | 3    |
| 2.1. Basic Principles.....   | 3    |
| 2.2. SMS History.....  | 3    |
| 2.3. SMS Network Elements.....                                       | 5    |
| 2.3.1 Base Transceiver System/Base Station Controller (BTS/BSC)..... | 5    |
| 2.3.2 Mobile Switching Center (MSC).....                             | 5    |
| 2.3.3 Signaling Transfer Point (STP).....                            | 6    |
| 2.3.4 Signaling System 7 (SS7).....                                  | 6    |
| 2.3.5 SS7 Application Protocols.....                                 | 6    |
| 2.3.6 SMS Protocols.....   | 7    |
| 2.3.6.1 Short Message Peer-to-Peer Protocol (SMPP).....              | 7    |
| 2.3.6.2 Universal Computer Protocol (UCP).....                       | 8    |
| 2.3.6.3 Computer Interface to Message Distribution (CIMD2).....      | 8    |
| 2.3.6.4 Open Interface Specification (OIS).....                      | 8    |
| 2.3.6.5 Telocator Alphanumeric Protocol (TAP).....                   | 8    |
| 3. Technical Analysis.....   | 9    |
| 3.1 Reliability.....   | 9    |
| 3.1.1 Definition of Reliability.....                                 | 9    |
| 3.1.2 Signaling Network Reliability.....                             | 10   |
| 3.2 Capacity.....  | 12   |
| 3.2.1 Wireless Network Capacity.....                                 | 13   |
| 3.2.2 Fixed Portion of the Wireless Network Capacity.....            | 21   |
| 3.3 Vulnerability Assessment.....                                    | 24   |
| 3.3.1 Denial of Service (DOS) Attack.....                            | 25   |
| 3.3.2 Service Interruption Attack.....                               | 25   |
| 3.3.3 Service Hijacking Attack.....                                  | 26   |
| 3.3.4 Spoofing.....  | 29   |
| 3.3.5 Non-Internet Based Methods of Attack.....                      | 29   |
| 3.4 SMS-Related Developments.....                                    | 30   |
| 3.4.1 Cell Broadcast.....  | 30   |
| 3.4.2 Enhanced and Multimedia Messaging.....                         | 32   |
| 3.4.3 Current and Future SMS Services.....                           | 35   |
| 4. Security.....   | 39   |
| 4.1 Application Originated Messages.....                             | 39   |
| 4.2 Mobile Originated Messages.....                                  | 40   |
| 4.3 Writer-to-Reader Security.....                                   | 42   |
| 5. Conclusions.....  | 45   |
| 6. Recommendations.....  | 47   |
| Appendix A - SMS Operation.....                                      | 49   |
| Appendix B - List of Acronyms.....                                   | 57   |
| Appendix C - List of References.....                                 | 61   |

## List of Figures

|   |    |
|---|----|
| Figure 2-1. Typical Handset-to-Handset Architecture.....                      | 5  |
| Figure 3-1. Network Elements for Routing and Processing SS7 Messages .....    | 10 |
| Figure 3-2: Sample 2-TRX Configuration.....                                   | 16 |
| Figure 3-3: Sample 4-TRX Configuration.....                                   | 16 |
| Figure 3-4: Sample 6-TRX Configuration.....                                   | 17 |
| Figure 3-5. Behavior of Performance Indicators in Congestion Situations ..... | 20 |
| Figure 3-6: Short Message .....   | 33 |

## List of Tables

|  |    |
|--|----|
| Table 2-1. SMS Messaging Protocols.....                  | 7  |
| Table 3-1. SMS and CBS Key Differentiators.....          | 30 |
| Table 3-2. Comparison of GSM Messaging Technologies..... | 32 |



## Executive Summary

The Short Message Service (SMS) allow textual messages to be delivered between SMS enabled, digital cell phones. These messages travel on the SS7 network in tandem with voice call signaling traffic. The service is extremely popular in Europe and Asia, generating tens of billions of messages a month. A number of factors slowed initial SMS acceptance in the United States, but traffic has grown remarkably in the last two years, and this trend is expected to continue.

SMS has also demonstrated utility as an alternative to voice communications. During the terrorist attacks on September 11, 2001, high traffic volumes made it extremely difficult to connect calls. However, SMS text messaging continued to operate and provided communications means for people who understood how to use it.

The conflux of these events raises the question: given that SMS shares the SS7 network, that popularity and awareness of SMS are increasing rapidly, and that SMS might be a person's means of communications during another crisis, how would the wireless network handle the surge in short messaging traffic? This Technical Information Bulletin (TIB) answered this question by researching multiple aspects of SMS technology, including wireless network reliability, capacity and congestion handling, security and vulnerability, priority services, and new SMS-related developments.

The TIB concluded that networks are theoretically capable of continued operation and service (including NS/EP service) during a crisis. However, it may be possible for extremely high volumes of SMS traffic, when combined with high numbers of voice call attempts, to interfere with a wireless network's performance. The absence of a definitive conclusion is due to the unavailability of key information: the network configuration details (known only by the network operator), the congestion handling algorithms (known only by the equipment manufacturers).

The report makes several recommendations:

- Characterize the anticipated traffic load (voice and messaging) better
- Obtain details of congestion handling algorithms from equipment vendors

Using this information, probable bottlenecks can be identified and a specific analysis can be performed for areas of concern.



# 1. Introduction

Short Message Service (SMS) is a relatively new feature in wireless telecommunications that allows short, textual messages to be delivered to cellular telephones. SMS is extremely popular in Europe and is gaining popularity in the US and worldwide. SMS messages are transmitted over the control network, Signaling System 7 (SS7), and not the bandwidth channels allotted to voice communications. SS7 is the basis for all control networks used by all major wireless and wireline telephone carriers. Disruption of SS7 operations could be devastating to the PSN and to NS/EP. SMS and related services, Enhanced Message Service (EMS) and Multimedia Messaging Service (MMS), are becoming as popular in the US as they are in Europe as evidenced by the introduction of digital picture and internet capable handsets. These additional features will greatly increase the resource requirements over SMS and, when combined with increased popularity, will impact the future load on SS7.

This Technical Information Bulletin (TIB) addresses two major SMS Markets; the United States and Europe. In Great Britain alone, over 68 million short messages were sent on Valentine's Day, 2003. In Europe, 10 billion messages are sent each month. In the United States, SMS traffic has increased tremendously to today where hundreds of millions of short messages are sent each month.

The goal of this TIB is to:

1. Perform an analysis of the vulnerability of SS7 to a huge increase in Short Messaging Service (SMS) traffic.
2. Examine the state-of-the-art of SMS applications in the PSN environment.
3. Examine the security aspects of employing SMS in an NS/EP environment.
4. Provide an analysis of future developments in SMS and transmission media other than PSN.

This TIB includes:

- A description of the SMS technology
- Capacity and load analysis for high traffic volume situations
- Security issues of using SMS in the PSN and other media.

Section 2 Background introduces SMS at a high level and describes the implementations used in North American wireless networks and also presents the protocols used in message delivery.

Section 3 Technical Analysis Investigation presents the different subjects that make up this TIB and include:

- The reliability of SMS delivery methods is discussed.
- The capacity and congestion behavior of the air interface and fixed network are presented.
- Security and vulnerability assessments of SMS are provided.

- The impact of several SMS-related developments, such as Cell Broadcast, EMS, and MMS is discussed.

Section 4 deals with the Security aspects of SMS including privacy over the airwaves and security as a message travels over SS7.

Section 5 Conclusions and Section 6 Recommendations provide a summary of the technical results and topics for further exploration and study.

## **2. Background**

### **2.1. Basic Principles**

SMS allows users to send and receive short textual messages directly on their cellular or PCS cell phones. SMS utilizes the handset keypad and menus to write, format, send and receive text. SMS messages can be sent from a PC over the internet, from other handsets, and from other internet capable, cellular devices. SMS is a digital service and as such is only available in areas where carriers have digital coverage; SMS messages will not be available in analog-only territories. Details of the SMS message transfer process is provided in Appendix A.

### **2.2. SMS History**

SMS was an unintentional success in the cellular industry. It had little promotion by wireless network operators and its early acceptance began in the European markets, where the youth segment started the trend. Several factors were responsible for this early European adoption.

- The tariff structure in Europe makes wireless voice calls expensive as compared to the U.S.
- Europe introduced pre-pay cellular service tariffs where people could pay for their airtime in advance.
- Manufacturers began making cheap, stylish, light weight, cell phones that had the SMS feature. This made cell phones very attractive to the European youth market.
- The European network operators initially were unable bill for SMS usage in pre-paid plans. The younger generation identified and was able to exploit this loophole, and the SMS industry was formed.

When the carriers figured how to bill pre-pay card customer for SMS usage, its usage slipped but picked up again after a few months as the service had become part of the culture. SMS was still cheaper than voice calls and thus deemed valuable enough to warrant paying for it.

As SMS penetration and revenues increased, wireless network advertising moved from business orientated applications to a more youth-segmented market. Asia was the next market but lagged the European markets by approximately two years with strong uptake and viral effect of the services, having at this writing mushroomed into the largest new wireless business. The U.S. is now following Europe and Asia with strong penetration of SMS and of its EMS and MMS variants.

The United States has lagged the remainder of the world in SMS usage and pervasiveness. Some reasons for this lag include the variety of wireless technologies used in the U.S., the technologies' varying implementation of 2-way SMS support, late support for inter-carrier messaging, little early marketing, and different pricing models. As an example of pricing, as of mid 2003, Verizon's America's Choice<sup>SM</sup> plan includes 300 peak time minutes and 3,200 off-peak minutes for voice calls for \$34.99 per month. The same plan allows SMS messages at \$.10 per message sent and \$.02 per messages received. In this example, voice calls are considerably less expensive than SMSs. Assuming a voice conversation is 3 minutes long, one could make

1100 of them for \$34.99 and assuming an SMS conversation (1 sent & 1 received) costs \$.12, one could only have 291. Voice communication in the US is considerably cheaper than SMS.

Until recently, it was difficult for subscribers on one wireless network to send messages to subscribers on other wireless networks. This was a deterrent to increased SMS usage. These difficulties were characterized by:

- Different allowable message sizes
- Separate 2-way messaging technologies (some operators use Wireless Application Protocol (WAP), not SMS, for 2-way messaging)
- Multiple character set conversion rules
- Inconsistencies in message priority definitions

In April 2002, the major carriers signed interoperability agreements and launched services to route short messages between networks. Several vendors, including Inphomatch, Telecommunications Systems (TCS), and Mobilespring, provide this functionality. Now virtually all subscribers on national wireless networks are able to send messages to each other.

In addition to the six national Tier I<sup>1</sup> wireless operators, there are over 70 Tier II carriers in the U.S., ranging from Alltel with over 7 million subscribers to small “mom and pop” carriers with less than 10,000 subscribers. The major carriers have interoperation agreements, whereas the smaller carriers are either examining SMS or have just launched SMS service. For the smaller carriers, purchasing and operating their own SMS Center (SMSC) is cost prohibitive. To solve the cost issue, many outsource their operations to one of a handful of companies providing SMSC connectivity. Companies offering this service include Numerex and TSI in the U.S.

During the 9/11 crisis voice calls in New York could not get through due to call volume, spectral limitations for cell phones, and damage to the local wireline networks. Voice calls in Washington DC were similarly affected in terms of volume, incoming call-blocking by the carriers, and spectral limitations. However, text messages in both locations were able reach cell phones [1, 2]. This demonstrates the potential for SMS to support the Government during National Security and Emergency Preparedness (NS/EP) situations.

---

<sup>1</sup> Tier I carriers have over 50 Points of Presence (POPs), are managed through a 24X7 Network Operations Center, have the ability to reroute and fall back to the PSN, and have redundancy in terminating locations. Tier II are similar with significantly fewer capabilities and presence [41]

### 2.3. SMS Network Elements

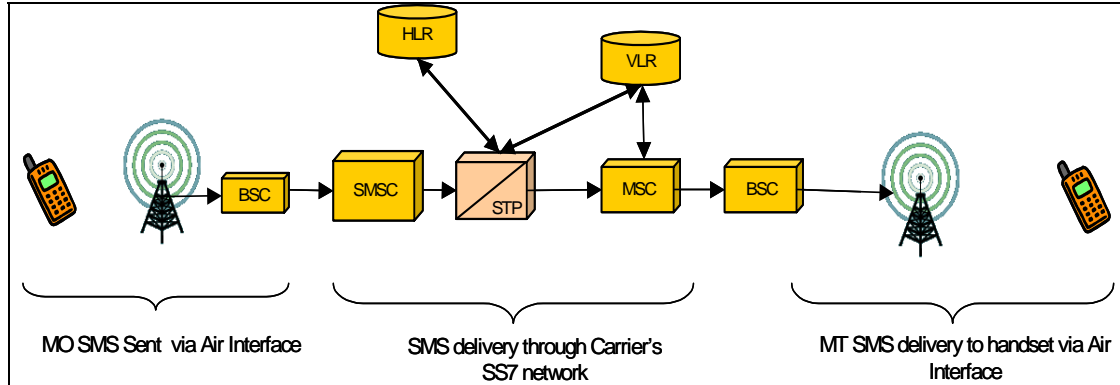


Figure 2-1. Typical Handset-to-Handset Architecture

#### 2.3.1 Base Transceiver System/Base Station Controller (BTS/BSC)

The Base Transceiver System (BTS), is the link between the Mobile Switching Center (MSC) and the wireless-to-handset interface. The BTS holds the radio infrastructure used to transmit the digital signals (control information and messages) and voice traffic to the cell phone. The BTS typically covers a radius surrounding a cellular transmission tower and is referred to as a macro-cell. Conversely, micro-cells are typically on buildings, in tunnels, in malls, or in arenas, and serve very specific locations. In many cases there is an intervening piece of equipment between the MSC and the BTS, it is called a Base Station Controller (BSC). The BSC controls a small region of cell sites, transcoding and routing voice, control, and messages to the various BTS towers located within the region.

#### 2.3.2 Mobile Switching Center (MSC)

The MSC is the main network element connecting the PSTN to the cellular phone system in a particular market. It manages, routes, and switches all traffic into and out of the cellular system. Generally each MSC connects to multiple BSCs, which in turn connect to multiple BTSs. Generally a switch can handle up to 200 +/- BTSs simultaneously. Aside from call setup and teardown, the MSC controls the authentication and processing of features for each subscriber. To send messages to or from any wireless handset, the carrier must possess a SMSC in their network. The SMSC acts as a centralized store-and-forward device that accepts messages and buffers or retains those messages until a suitable deliver time (i.e., the cell phone is powered on and the location known). There are a number of SMSC manufacturers. Several of the popular manufacturers within the U.S. include LogicaCMG, TCS, Comverse, Ericsson, SchlumbergerSema, and Motorola.

The Home Location Register (HLR) and Visitor Location Register (VLR) are databases within the MSC (HLR/VLR), or in some cases external to the MSC Standalone-HLR (SHLR), that house all the following data:

- Subscriber information

- Location
- Feature set
- Phone status (on or off)
- SMS location status (i.e., if the phone is in an SMS-capable location)
- Additional information essential to making a wireless network function

The HLR is a subscriber profile database maintained the cellular provider of record. It is used for user authentication and contains subscriber data related features and services. The VLR is the subscriber profile database allowing feature and user authentication as well as critical features for routing messages and subscriber information to the SMSC. When a cellular user moves from one MSC to another, the VLR entry is torn down and re-entered (registered) in the new MSC VLR via SS7 connections.

### **2.3.3 Signaling Transfer Point (STP)**

STPs are routers in an SS7 network and are analogous to routers in an IP environment and provide routing of messages, call setup, handoff, address translation, and hot-standby functionality. STPs are universally run in pairs, generally geographically diversified to provide reliable and redundant, carrier class networks. Any STP within the SS7 network can be accessed by any other STP within the network thus providing for mission-critical systems connectivity.

As the volume of SS7 messages increases, carriers looking to reduce costs are turning to SS7 over IP solutions, which replace traditional STPs and link sets with an IP backbone.

### **2.3.4 Signaling System 7 (SS7)**

Signaling System 7 (SS7) is the transport element for SMS traffic. It is the network control protocol for telephone service providers worldwide. SS7 consists of four levels that map, in principle, to the OSI model. The primary use of the SS7 system is to control the network. It is used to transport data for both landline and wireless telephone systems and has become the standard for signaling worldwide. Messages in SS7 travel independently of voice traffic, from one network to another, riding over packets. The protocol is used in both wireless and wireline networks, for such things as call setup and teardown, routing for 800 numbers and 900 numbers, database lookup, and in wireless networks for querying the HLR via IS-41 for registration, location, feature provisioning, authentication, caller ID, and SMS messaging delivery.

### **2.3.5 SS7 Application Protocols**

#### **2.3.5.1 IS-41 – Personal Communications Network (PCN) to PCN Intersystem Operations**

The PCN to PCN Intersystem Operations (IS-41, aka ANSI-41) standard in the U.S. was originally set up for cell phone users to be connected while moving from one calling area to another. Formerly, the user would arrive in a new calling area and have to call a number or dial a code specific to that operator and register themselves in the new network. This worked fine for long term stays, but it was cumbersome for callers traveling through an area. Now when a caller



moves to a new location the IS-41 system provides for seamless roaming, call and message delivery, and the intersystem operation of cellular networks such as analog Advanced Mobile Phone System (AMPS) ANSI-136 or digital Code Division Multiple Access (CDMA) IS-95 networks. The intersystem support provided by IS-41 includes automatic roaming, intersystem handoff, intersystem operation, administration, and maintenance. ANSI-41 defines the interfaces between MSCs and between the MSC and the SMSC.

### 2.3.5.2 Global Mobile System (GSM) Mobile Application Part (MAP)

GSM MAP serves the same purposes as IS-41 and communicates with similar network elements such as the MSC, HLR, and VLR, but it was designed for GSM and to support SMS. The technology is essentially the equivalent to IS-41.

### 2.3.6 SMS Protocols

The protocols used by applications to send and receive messages are identified in Table 2-1.

**Table 2-1. SMS Messaging Protocols**

| Protocol   | Owner/Creator                     |
|--|-----------------------------------|
| SMPP (Short Message Peer to Peer, P2P)             | SMS Forum / Logica                |
| UCP (Universal Computer Protocol)                  | CMG (now LogicaCMG)               |
| CIMD2 (Computer Interface to Message Distribution) | Nokia                             |
| OIS (Open Interface Specification)                 | Sema Group (now SchlumbergerSema) |
| TAP (Telocator Alphanumeric Protocol)              | PCIA                              |

#### 2.3.6.1 Short Message Peer-to-Peer Protocol (SMPP)

SMPP is the predominant protocol used for SMS today. The most widely used version is SMPP V3.4, with V5.0 being introduced in 2002 and most likely phased in over the years 2003 through 2005. SMPP-V4 does exist, but it is a proprietary version developed for application in the Japanese Personal Digital Communications (PDC) market. It was devised to meet with specific needs of the customer. V3.4 represents the official enhanced SMPP protocol and is currently being migrated to SMPP-V5.0. SMPP-V3.4 is the defacto standard for the following reasons:

- Widely deployed
- Air interface independency
- Independent standards body ownership (SMS Forum)
- Flexibility
- Available of Software Toolkits
- TCP/IP-based
- The source is not required to know the destination
- Able to send based on dialable (NPA-NXXX) number

There are a number of enhancements within SMPP-V5.0, including the following:

- Additional routing elements
- Source identification details
- Destination identification details
- Point-to-point connections required
- Number portability
- Global Unique Message ID

### **2.3.6.2 Universal Computer Protocol (UCP)**

The Universal Computer Protocol was developed by ETSI (European Telecommunications Standards Institutes). It is the leading protocol in Europe and is used extensively throughout all the countries and carriers. This protocol must be considered when evaluating SMS usage and delivery schemes. The most widely used version is UCP-V3.5, which is supported by most SMSC vendors. UCP-V4.0 was released in 2001 and is in the adoption process.

### **2.3.6.3 Computer Interface to Message Distribution (CIMD2)**

Computer Interface to Message Distribution (CIMD2) was developed by Nokia. It is a protocol that is used to connect applications to a messaging platform. Applications are defined as programs that are capable of sending and receiving messages. The main purpose of a CIMD2 connection is to transfer messages from the application to a cell phone and from the cell phone to the application. Other kinds of information can also be conveyed over the interconnection (e.g., status reports from the GSM network to the applications). The CIMD2 protocol supports TCP/IP sockets, X.25 packet assembler/disassembler (PAD), and serial ports (modems). Since Nokia and CMG are the only manufacturers supporting CIMD2 on their SMSC platforms, most manufacturers considered it less important to implement CIMD2 [3].

### **2.3.6.4 Open Interface Specification (OIS)**

The Open Interface Specification (OIS) was developed by the Sema Group, a French company that was acquired by Schlumberger. This protocol is not as widely used as SMPP or UCP, but it is used in select locations in Europe and extensively by Vodafone. OIS-V5.1 is the most common version in use.

### **2.3.6.5 Telocator Alphanumeric Protocol (TAP)**

The Telocator Alphanumeric Protocol (TAP) was developed for the Personal Communications Industry Association (PCIA) for sending messages between paging companies and SMS providers in the United States. TAP is supported through dialup connections as a legacy product for many of the providers. It does not support many of today's SMS features such as ring tones, pictures, etc.

### 3. Technical Analysis

The technical analysis section presents the research results for several issues relevant to NS/EP. They include:

- What load would be placed on the wireless network by such usage during a crisis?
- How would the network respond?
- How would other services, such as voice calls, be affected?
- How secure are the messages?
- Are there clear vulnerabilities to be aware of?

Significant attention is given to reliability, capacity, security and vulnerability assessments. The section concludes with a short discussion of other relevant, SMS-related developments.

- Cell broadcast
- Enhanced messaging
- Multimedia messaging
- Current and future SMS services

#### 3.1 Reliability

This section examines the reliability of the signaling network and addresses the communications protocols that allow reliable delivery of messages at the signaling network level as well as at the overarching level of SMS message transfer. These aspects of network reliability will be addressed by grouping network components and information transfer processes into the following two discussions:

- **Signaling Network Reliability** – how a signaling network obtains reliability through the physical paths messages take, and the protocols for accurately transferring messages across those paths.
- **SMS Transfer Reliability** – how methods for reliable message delivery are adjusted for factors such as cell phone status or ability to reach the destination.

##### 3.1.1 Definition of Reliability

Network reliability will be defined for this study as union of the concepts of *availability*, *redundancy*, and *fault detection*. These three concepts are interrelated as follows:

- A network with high availability will have little or no downtime (high uptime).
- Redundant network components usually results in minimal downtime.
- For redundant network components to contribute to reliability, fault detection must be available.

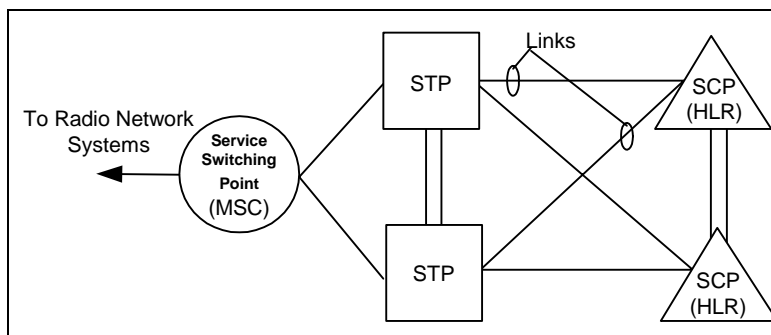
Often, the term fault tolerant is used when discussing reliability. A fault tolerant system can be equated with a completely reliable system where high availability is achieved by having redundant elements that can provide ongoing service during fault conditions. Even though a system may be declared “fault tolerant”, or “100 percent reliable,” fault detection and resolution typically depend on human interaction especially in the fault resolution process, which must occur within an appropriate timeframe so that redundancy is not compromised.

With these concepts in mind, this analysis will examine components and processes within a generic wireless communication system as they relate to transport of SMS messages.

### 3.1.2 Signaling Network Reliability

The signaling network portion of any telecommunications network is crucial to passing any form of subscriber traffic. If any signaling message does not reach its destination, phone calls may not be completed or may be disconnected, and SMS messages might not reach their destinations. A telecommunications signaling network is therefore always designed and deployed with extremely high reliability in mind.

A modern signaling network, based on SS7, encompasses parameters such as physical connectivity (e.g., DS0 links), types of messages, and protocols (e.g., link status messages within the Message Transfer Part protocol). Also implied are fixed elements of the network for routing and processing SS7 messages, as shown in Figure 3-1 below. This section will describe how these network elements offer highly reliable transport of signaling and SMS traffic.



**Figure 3-1. Network Elements for Routing and Processing SS7 Messages**

#### 3.1.2.1 Signaling Network Components

An SS7 network used within a wireless network has the following core components [4]:

- **Links** – the DS0, and, more recently, DS1 circuits between any two SS7 network elements.
- **Signaling Transfer Points (STPs)** – routing devices that send incoming SS7 messages to their destination by evaluating address information in each message.

- **Service Control Points (SCPs)** – databases that are queried by other SS7 network elements that need call and message routing information in real time. An HLR is the most common SCP in a wireless network.
- **Service Switching Points (SSPs)** – network elements, such as the MSC, that utilize SS7 messaging to provide the connectivity requested by a user's action. An example of such an action is a wireless subscriber dialing a phone number causing an SS7 transaction to be originated from an MSC.

All of these SS7 components offer high reliability based on the following network architecture elements:

- Signaling Transfer points are highly available computer systems, using the same platform redundancy, software redundancy, and fault detection mechanisms described with respect to MSCs.
- Links are provisioned in pairs from any network node to any other network node. In the diagram above, the MSC has a pair of links (one to each STP), and the STPs have a pair of links to each other.
- Each individual link in a pair of links travels a separate geographic route.
- At each building housing an SS7 network node (including MSCs and HLRs), link pairs are brought in through a separate building entry point. This helps prevent outages such as those caused by backhoe and drilling accidents.
- Each member of a link pair is attached to a port on an SS7 network node under separate processing control and separate physical location from the other link member.

### 3.1.2.2 Signaling Path Reliability

In addition to the signaling path reliability provided by designing circuit redundancy and geographically diverse circuit paths, SS7 networks utilize protocols that ensure path availability and integrity through; standardized physical and link definitions, message sequencing, error detection and correction, and automatic network management processes. These functions equate to levels 1 through 3 of the SS7 protocol stack and form the basic layer of SS7 networks known as Message Transfer Part (MTP). The SS7 protocol also provides the following functionality:

- Load sharing across redundant links - An SS7 network node will distribute traffic equally across all links. Through this load-sharing scheme, an outage on a link will not lead to any delays caused by failover to a standby link, or any abrupt cancellation of transactions based on messages sent through the link.
- Link loading and load handling mechanisms - A typical SS7 network employs two-way redundancy (links and STPs are provisioned in pairs). With this in mind, link loading is kept under 40% utilization by the wireless carrier. This design allows an outage of one half of a pair, causing traffic on the other half during the outage to be kept under 80% utilization and allowing 20% of its capacity available for traffic peaks. If traffic exceeds 100% congestion, handling procedures at each end will handle the load by: buffering, processing only messages with high priority, or the discarding of messages.

- Link monitoring and status handling - Regular link status checking is conducted throughout an SS7 network. If a link becomes unavailable, messages are sent to ensure both attached nodes are aware of the outage, and automatic link test and link restart procedures are begun. Unless the link restart procedure is successful, a link will remain unavailable for traffic at both ends until the problem is resolved.
- Route redundancy - Whereas links are physical paths between nodes, routes comprise multiple links that together takes messages from one node to another and does not require a direct link connection. In a network with multiple pairs of STPs, multiple routes to a destination become available, so that failure of a link somewhere in the path does not impede message delivery.
- Route monitoring and status handling - When routes become unavailable due to outages or maintenance, traffic management messages are sent to affected nodes, or network elements that utilize a particular route between two nodes. These nodes are made aware that they must send traffic across a redundant route. When an unavailable route returns to service, traffic management messages are sent again to notify all concerned nodes about the availability.

### **3.1.3 SMS Message Transfer Reliability**

Message transfer reliability is provided through the SS7's Signaling Connection and Control Part (SCCP). While the MTP layer ensures that messages can be transported from node to node, SCCP provides end-to-end connectivity across multiple nodes and connection-oriented messaging and transaction-oriented message handling. SCCP information within SS7 messages provides addressing elements that allow messages to be routed according to various parameters such as normal 10 digit phone number or an SS7 point code and subsystem number<sup>2</sup>.

## **3.2 Capacity**

Because of the economics of designing, building, and maintaining very expensive telecommunications infrastructure, carriers dimension their networks to handle the normal, day-to-day traffic loads and not the absolute peak loads resulting from Mother's Day or Christmas calls. Elements in a network are dimensioned according to the busy hour load (the load placed on the network during the busiest part of a day). It is not cost-effective to dimension networks for high-traffic events that occur infrequently. High-volume events in specific areas, such as a sports events attended by a large crowds, are supported by installing additional, temporary capacity. Long-term growth is managed through constant monitoring of network metrics such as link utilization, processor utilization, and blocking rates.

---

<sup>2</sup> A point code is a representation of a network element's internal SS7 address. A subsystem number provides identification of the software process at the destination that is to receive the message.

SMS capacity analysis is complex for several reasons, including the following:

- Six wireless operators offer nationwide cellular service. Each operator's network is implemented differently.
- Operators do not use homogeneous technology nor follow the same air interface standards across their individual networks. Cingular and AT&T Wireless, for example, operate a mixture of Time Division Multiple Access (TDMA) and GSM networks.
- Traffic patterns are constantly evolving. Aside from the increased use of wireless services, there are several discernible patterns:
  - Due to its increasing popularity, SMS traffic is increasing at an accelerated rate and is altering the network load.
  - New services are causing a shift in network utilization. Some are SMS-based (e.g., Microsoft offers SMS versions of Hotmail and MSN Messenger). Others offer multimedia data services, such as Sprint PCS' *PCS Vision* or AT&T Wireless' *mMode*.
  - Operators will minimize capital expenditures for an older technology while spending money on, and migrating users to, the newer technology. This traffic shift is a constant dynamic and last for several years.
- Traffic analysis will require specific knowledge of each network. Each cellular network has multiple HLRs, VLRs and MSCs that are based on the geographical coverage area and are dimensioned for different capacities. These elements are connected by hundreds of STPs and thousands of SS7 links. Each node or link may operate at a different utilization. Finally, subscribers are mobile, causing traffic patterns to shift.

In order to make the capacity problem tractable, this TIB has limited its discussion to the GSM SMS implementation. GSM was chosen for the following reasons:

- The GSM standard has the broadest scope of all the wireless technologies. More aspects of the GSM system are specified than TDMA, CDMA or iDEN networks
- Specifications are publicly available
- All GSM networks offer 2-way SMS. Some wireless operators using other technologies do not offer 2-way SMS, or offer 1-way SMS augmented with WAP solutions, making analysis more difficult.

### **3.2.1 Wireless Network Capacity**

#### **3.2.1.1 Voice Traffic Load**

Historically, periods of crisis lead to increased load on telecommunications networks. The following statistics describe the loads that were placed on major networks during the September 11 tragedy [5]:

- AT&T long distance recorded 101 million more calls that day than the previous all-time high. Network carried on average 4 million calls every five minutes after the attacks, double the normal call volume [42]
- Verizon Wireless experienced traffic rates from 50 percent to 100 percent higher.
- Attempted calls via Cingular Wireless increased 400 percent in Washington and 1000 percent in New York.

The finite resources available in many wireless networks resulted in high blocking levels.

### 3.2.1.2 SMS Traffic Load

The following assumptions and scenarios are presented

- **ASSUMPTIONS**

To determine the load that SMS traffic will place on a network during a crisis, the SMS capacity on the air interface and the offered load should be considered. As a general rule, an SMS message occupies a Standalone Dedicated Control Channel (SDCCH) channel for 4-5 seconds. A sector with 4 Transmitter/Receiver (TRXs) might have 8 SDCCH channels and a capacity of nearly 7200 SMSs/hour, or 120 SMSs/minute.

Each message lasts 4 seconds  
 Each SDCCH can handle 15 messages/minute  
 Each sector has 8 SDCCHs, therefore -  
**Each sector can handle 8 X 15 or 120 Messages /minute**

The average length of an English word is 5.98 characters. Empirical measurements show the average speed at which users enter text on a cell phone is between 7.9 and 8 words per minute. Using these statistics, it can be calculated that a 50-character SMS would hold 8.3 words, and would take around than 60 seconds to enter. Therefore the following scenarios, which examine SMS usage in Washington, D.C., and Manhattan, will assume the average person can send 1 SMS per minute. The scenarios do not consider re-send attempts.

Average word is 5.98 characters  
 Average SMS message contains 50 Characters or 8.3 words, and  
 Average speed of SMS entry is 7.9 - 8.0 words/minute  
**Therefore, a user can enter one SMS in about a minute**

- **SCENARIO 1: Washington, DC<sup>3</sup>**

As of April 2000, Washington, DC, had a population of 572,059 distributed over 68.2 square miles [6]. This equates to an average density of 8,388 people/sq mi. Assuming Washington is covered by 40 cell sites with 120 sectors, each sector would cover an average of 1/2 to 3/4

---

<sup>3</sup> The above estimates are conservative. The District has an influx of workers during the day, and those workers have a higher cell phone penetration than the general population. Therefore, the load could be considerably higher.



square mile (there is overlap in the coverage areas). Based on coverage area and population density, an average sector covers approximately 6000 people. Washington DC has around 60% penetration [7], or 3600 subscribers per sector. They could generate 3600 messages/minute in each sector, or 30 times greater than the 120 SMSs/min a sector can process as calculated above.

- **SCENARIO 2: Manhattan, New York**

As of 2000, Manhattan had a population of 1,318,000 [8] distributed over approximately 31.1 square miles [9]. This leads to an average density of 42,379 people/sq mi, approximately 5 times the density of Washington, DC. Assuming a similar sector coverage area as presented in Scenario 1 yields an approximate maximum of 30,000 people covered by each sector. Using similar cell phone penetration rates, 18,000 subscribers are covered by each sector. If these users generate 18,000 messages/minute in each sector equates to 150 times the SMS capacity available. Even the highest-capacity sectors, using 12 SDCCH channels, would need 100 times more capacity to meet this load.

- **CONCLUSIONS**

By examining the Washington, DC, and Manhattan scenarios, it can be concluded that, if SMS were used extensively during a crisis, a significant SMS load could be placed on a network. Individually, the voice load and SMS load are multiple times higher than the engineered capacity at each sector. This analysis has not considered several factors that might increase load, such as messages originating from other sources (e.g., the Internet) and terminating in the congested area. It has also not considered message re-send attempts after failures, which add to network load [10].

### **3.2.1.2.1 Air Interface Capacity**

Another issue is the effect of high SMS traffic volumes as messages traverse the core SS7 network. This section investigates network performance when traffic volumes *significantly exceed* the wireless network's designed capacity.

The potential for delivery process bottlenecks exists at several points in the air interface. This section discusses the capacity limits of each element of this interface and describes what happens when these limits are reached.

In GSM networks, there are four potential congestion areas:

- Traffic Channel (TCH) Congestion
- Standalone Dedicated Control Channel (SDCCH) Congestion
- Random Access Channel (RACH) Congestion
- Paging Channel (PCH) Congestion

### 3.2.1.2.1.1 Traffic Channel (TCH) Congestion

TCH congestion is the most common type of congestion and occurs when all TCHs in a sector are occupied with existing calls. This leaves no capacity for new conversations, so all new voice calls will be blocked. This congestion is common because traffic channels are a finite resource and are occupied for a relatively long period of time (i.e., the duration of a conversation).

In GSM, the number of TCHs available is a function of the number of TRXs in each sector and the assignment of time slots in each TRX. The number of TRXs in each sector is a function of the network's frequency plan. In many lightly populated areas, it is common to have 2 TRXs in each sector. A typical time slot assignment, as illustrated in Figure 3-2, might be as follows:

|       | 1                                       | 2   | 3   | 4   | 5   | 6   | 7   | 8   |
|-------|---|-----|-----|-----|-----|-----|-----|-----|
| TRX 1 | BCCH, FCCH,<br>SCH, AGCH,<br>PCH, SDCCH | TCH | TCH | TCH | TCH | TCH | TCH | TCH |
| TRX 2 | TCH                                     | TCH | TCH | TCH | TCH | TCH | TCH | TCH |

*One time slot is dedicated to control channels and 15 are dedicated to voice channels. The wireless operator chooses the actual distribution of time slots between control (orange) and traffic (green) channels.*

**Figure 3-2: Sample 2-TRX Configuration**

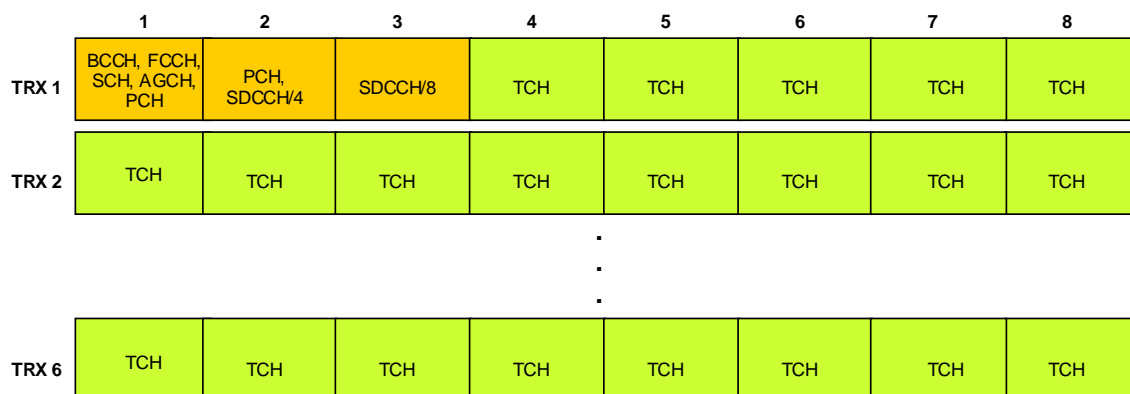
In urban areas, it is common for 4 TRXs to be available in each sector. A typical time slot assignment, as illustrated in Figure 3-3, might be as follows:

|       | 1                                | 2       | 3   | 4   | 5   | 6   | 7   | 8   |
|-------|----------------------------------|---------|-----|-----|-----|-----|-----|-----|
| TRX 1 | BCCH, FCCH,<br>SCH, AGCH,<br>PCH | SDCCH/8 | TCH | TCH | TCH | TCH | TCH | TCH |
| TRX 2 | TCH                              | TCH     | TCH | TCH | TCH | TCH | TCH | TCH |
| TRX 3 | TCH                              | TCH     | TCH | TCH | TCH | TCH | TCH | TCH |
| TRX 4 | TCH                              | TCH     | TCH | TCH | TCH | TCH | TCH | TCH |

*Two time slots are dedicated to control channels; in this example, the second time slot is used for 8 SDCCH channels. 30 time slots are dedicated to voice channels. The wireless operator chooses the distribution of time slots between control (orange) and traffic (green) channels.*

**Figure 3-3: Sample 4-TRX Configuration**

In order to provide necessary capacity, networks in dense urban areas may have 6 TRXs on each sector. These are usually organized as shown in Figure 3-4:



*Three time slots are dedicated to control channels; the second and third time slots provide 12 SDCCH channels and additional paging channels. 45 time slots are dedicated to voice channels. The wireless operator chooses the distribution of time slots between control (orange) and traffic (green) channels.*

**Figure 3-4: Sample 6-TRX Configuration**

The best way to provide additional voice channels is to provide more TRXs at each sector. A more aggressive frequency plan can provide this, but will increase interference, leading to poor voice quality and more dropped calls. In addition, modifying a system’s frequency plan is a time-consuming engineering process, which is a deterrent. NCS has worked with wireless operators to circumvent this problem. The result is the Wireless Priority Service (WPS), which provides select individuals higher priority access to TCHs.

### 3.2.1.2.1.2 Standalone Dedicated Control Channel (SDCCH) Congestion

Before a voice conversation begins or an SMS is sent, a number of operations are performed between the cell phone and the network. These include the following:

- Authenticating the handset
- Defining and turning on encryption
- Assigning the cell phone a new Temporary Mobile Station Identifiers (TMSI) (this provides increased anonymity for the cell phone user)
- Sending the dialed digits (or SMS payload) from the handset to the network

These operations are performed on the SDCCH. The actual SDCCH data rate is quite low, 782 bps [11].

SDCCH bursts carry 23 bytes, and as many as 8 bursts may be required to carry the payload of a single SMS message (not including the bursts required for authentication, encryption, and TMSI

assignment). Once these messages are accounted for, generally 4-5 seconds are required to transmit an SMS across the SDCCH.

As described above, a typical sector configuration contains 4 SDCCHs. A higher-capacity sector with more TRXs might contain 12 SDCCHs. Assuming no other traffic, a sector with 4 SDCCHs could transmit short messages at the rate of 3600 messages per hour. A sector with 12 SDCCHs could transmit short messages at a rate of 10,800 messages per hour.

Due to the relatively small number of SDCCH channels, their low capacity, and their relatively long hold times, it is possible for SDCCH channels in a sector to become congested. If the SDCCH channels in a sector become busy and none are free, new calls and short messages will fail, regardless of whether they are cell phone originated or cell phone terminated.

More SDCCH channels can be provisioned in each sector to prevent SDCCH congestion. Unfortunately, the air interface resources in each sector, which are a function of the number of TRXs, is constant. In order to dedicate more time slots to SDCCHs, the number of timeslots for TCHs must be reduced, decreasing the amount of voice traffic a sector can support. GSM engineers are forced to make a tradeoff - they can provision more signaling channels, or they can provision more voice channels, but they cannot do both.

#### **3.2.1.2.1.3 Random Access Channel (RACH) Congestion**

In order to obtain an SDCCH, cell phone must ask the sector for a channel assignment. This is done by sending an access burst on the RACH. This burst contains only a few bits of information, including the type of operation the cell phone wants to perform [12] listed below:

- Mobile originating call establishment or packet mode connection establishment
- Emergency call establishment
- Short message service
- Supplementary service activation

If the sector has an SDCCH available, it will be assigned to the cell phone.

The RACH is shared by all the cell phone in a sector. To coordinate access to the RACH, cell phone use a Slotted ALOHA protocol<sup>4</sup>. The GSM design engineer has a few tools to improve RACH performance:

- 1) The number of frames dedicated to RACH can be altered. However, this would require either
  - a. A reduction in SDCCH frames, or
  - b. A reduction in available TCHs
- 2) Back off and retry parameters can be adjusted.

---

<sup>4</sup> Slotted ALOHA Protocol is an access control technique for multiple-control access transmission media. It uses the same packet structure but in well defined time slots

Ultimately, however, the random access of wireless channels is a difficult problem and solutions are very constrained. Slotted Aloha works well for small numbers of users but degrades quickly as more users desire access. Its peak theoretical throughput is only 0.368 of the random access channel capacity. Performance degrades with increasing load and ultimately goes to zero [13].

#### **3.2.1.2.1.4 Paging Channel (PCH) Congestion**

The traffic carried on paging channels in each sector is an aggregate of all mobile terminal (MT) (mobile terminals are cell phones, pagers, PDAs etc.) traffic in the location area. Researchers study paging algorithms for potential congestion mitigation because cell phone terminated call setup and SMS paging requests are broadcast on every PCH in a location area. Location areas may contain dozens of cell sites, so this aggregate paging load can become significant. If the paging channels in a sector become congested for too long, subscribers will not be able to receive calls or short messages.

A GSM design engineer can improve PCH performance by increasing the number of frames dedicated to PCH. However, this requires either:

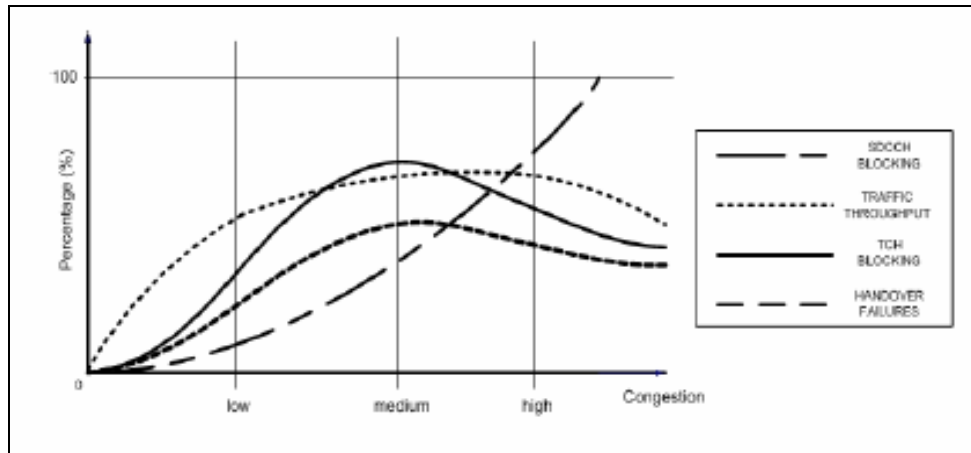
- 1) A reduction in SDCCH frames, or
- 2) A reduction in available TCHs

#### **3.2.1.2.1.5 Congestion Likelihood**

The performance of GSM networks heavily depends upon the availability of the channels described above. To track performance, a network provider collects performance measurements which act as metrics to indicate how well the network is performing. Some important metrics that describe congestion include the following:

- **Call Setup Success Rate:** the ratio of successful call setups to unsuccessful call setups.
- **Handover Success Rate:** the ratio of successful handovers to unsuccessful handovers.
- **SDCCH Blocking Rate:** the ratio of SDCCH requests that were not served to the total number of SDCCH requests.
- **TCH Blocking Rate:** the ratio of TCH requests that were not served to the total number of TCH requests.

Researchers have analyzed GSM measurements from operational networks to understand how networks react to different types of congestion [14]. Figure 3-5 below shows air interface performance as load increases through low, medium, and high levels. For the purpose of the study, high congestion was defined as blocking rates of over 5% for SDCCH or RACH, or blocking rates of over 10% for TCH.



**Figure 3-5. Behavior of Performance Indicators in Congestion Situations**

Network throughput, defined as the traffic that was successfully processed, increased as the load increased from low to medium levels. However, as congestion reached high levels, the total throughput actually began to decrease. This was caused by congestion in resources required to successfully complete calls, such as SDCCH channels. SDCCH blocking increased continuously as load on the network increased. This is to be expected, as only a fixed number of SDCCH channels are available to handle an ever-increasing load.

TCH blocking increased through low and medium levels of congestion. However, an interesting phenomenon occurs as the load increases: the TCH blocking actually begins to decrease. This is due to an increase in SDCCH blocking; as SDCCH channels become congested, cell phones are not able to complete the signaling steps required to request a traffic channel, reducing the total number of requests for traffic channels. At the same time, some users will complete their calls normally, thus freeing some previously used traffic channels and reducing traffic channel blocking. A similar effect is seen with handoff failures. While this may seem a benefit, it is important to remember that as traffic channel congestion decreased, the total number of calls serviced also decreased due to the lack of SDCCH channels. What is not shown in Figure 3-5 is as load increases even further, cell phones cannot reach the network. There may also be increased blocking on the RACH channel due to the Slotted Aloha algorithm. Blocking on the forward link was also considered. While the PCH might appear a more likely candidate for congestion than, say, the RACH, it benefits from scheduling improvements not available to the RACH.

In summary, the following types of congestion are expected as network load increases:

- 1) Calls begin to fail because traffic channels, TCHs, are blocked. SDCCH channels are often still available, however. This means that SMS can continue operating even when no more traffic channels are available.
- 2) As load increases past this point, SDCCH channels begin to block. This causes voice calls and short messages to fail.

- 3) At very high levels, cell phones may not even reach an SDCCH channel due to RACH congestion, PCH congestion. This also causes voice calls and short messages to fail.

### **3.2.2 Fixed Portion of the Wireless Network Capacity**

#### **3.2.2.1 BTS and BSC**

The GSM subsystem consists of two components, the base station (BTS) and the base station controller (BSC). In North American networks, these two components are not separated and are referred to collectively as the “base station”. Some of the functions that the BSC performs are placed in the base station, and some are performed at the switch, or MSC. In the GSM system, the BTS is connected to the BSC through E1 lines (T1 lines in the U.S.). The protocol used on this connection is LAPD (Link Access Procedure, D-Channel), the same protocol used on the ISDN D-channel for call setup and signaling. SS7 is not used on this link.

There are three types of traffic carried on the link, including Radio Signaling Link (RSL), Operation and Maintenance Link (OML), and Layer 2 Management Link (L2ML) [15]. Short messages are carried inside the RSL slots. In the event of congestion the system gives priority to OML and L2ML traffic. The BSC implements a priority or flow control algorithm so that if the processor utilization goes to high in BTS or BSC, or if some of the control channels or access channels become congested, then the controller will begin to reduce the amount of traffic it handles.

When the overload condition is sensed, the BSC will reduce the amount of traffic it supports. If the overload condition continues, the amount of traffic supported will be reduced again. This is referred to as a step down algorithm. This algorithm continues as long as the overload condition is present. When a timer expires and the overload condition is no longer present, the controller will step up the amount of traffic it supports until the traffic level returns to normal and the overload condition disappears [16].

It is possible that some of the traffic will be dropped. If the base station itself is overloaded, it will begin to reject messages sent by the BSC. For example, if the BSC tries to assign a traffic channel to a cell phone, the base station will delete the message. It is difficult to predict which messages will be dropped and which will be kept. For example, voice messages may have priority over short messages. If the system is overloaded, the final implementation is dependent on the vendor and is outside the scope of this TIB.

In the case of an overload, the BSC is aware of the overload and of the type of operation the cell phone is requesting, e.g., an SDCCH to make a phone call, send a short message, or make an emergency call. It’s conceivable that as part of the flow control algorithm, voice requests may take priority over SMS requests; emergency calls may take priority over regular voice calls, etc. This prioritization is dependent upon the vendor implementation.

### 3.2.2.2 BSS-MSC Interface

The connection between the base station subsystem (BSS) and the MSC is called the A-interface. The A-interface uses SS7 for message transfer. This means that Message Transfer Part, Level 1 (MTP1), Message Transfer Part, Level 2 (MTP2), Message Transfer Part, Level 3 (MTP3), and Signaling Connection Control Part (SCCP) provide connectionless and connection-oriented SS7 transport for the application part.

In the case of the A-interface, BSS Application Part (BSSAP) is the application that runs on top of SCCP. BSSAP has two subparts, called Direct Transfer Application sub-Part (DTAP) and BSS Management Application Sub-Part (BSSMAP). DTAP is used to send messages from the MSC directly to the cell phone by mapping the message on the air interface channel and transmitting it to the cell phone (not interpreted by BSS). BSSMAP is used to send messages from the MSC to the BSS. These messages are interpreted by the BSS and are used for signaling, connection setup, etc.

Every signaling channel on the air interface has an analogous SCCP connection on the A-interface. Therefore, when a cell phone wants to send an SMS, an SCCP connection is created on the A-interface. BSSMAP is used to setup the channel, while DTAP carries messages directly to and from the handset. According to the specification, no special flow control or congestion algorithms are defined on the A-interface. In the event of congestion, standard SS7 congestion routines employed (see SS7 section). Thresholds and abatement procedures are system-dependent [17].

### 3.2.2.3 MSC

The most common MSC loading problems concern processor utilization. As the number of messages increases, more CPU cycles are required to process them. As CPU utilization on the MSC approaches a high value (e.g., one threshold may be 80%, another threshold may be 90%), traffic to the MSC must be reduced.

The MSC can indicate high CPU utilization by sending an OVERLOAD message to the BSS. How the BSS handles this message is implementation-dependent; one suggested method of relieving congestion is to throttle back new connection attempts (e.g., cell phone originated voice and SMS calls). By employing access control features on the RACH [18], the BSS can step up/step down cell phone originated attempts, usually by blocking it in increments of 10%. As the overload condition dissipates, the BSS can gradually increase access. During this MSC congestion period, some users would not be able to send short messages or make phone calls. This algorithm does not affect emergency calls [19].

### 3.2.2.4 SS7

SS7 transports signaling traffic between MSCs, HLRs, VLRs, and SMSCs. GSM networks also use SS7 to carry signaling information between the BSS and MSC (the A-interface). This interface is discussed separately because it employs a different user part. North American networks employ a cell phone application part called IS-41. GSM networks employ another



application part, called GSM MAP. Both application parts use SS7's TCAP layer and enable signaling traffic, such as call completion, mobility management, and short messaging, required by wireless networks.

Congestion-handling mechanisms are built into SS7 networks. The philosophy of SS7 congestion handling is to throttle traffic as close to the source as possible [20]. This serves multiple purposes:

- By throttling traffic at the edge of the network, fewer network resources are used. Calls and/or messages are halted before many resources are assigned to them.
- Signaling requirements for new traffic are usually greater than existing traffic. By preventing new calls from taking place, the SS7 network hopes to alleviate the subsequent load placed by a call's (or message's) signaling traffic.
- The edge of the network is closest, literally and logically, to the end user. Therefore, it is best positioned to make intelligent decisions about how traffic should be throttled.

This congestion approach is designed to keep the network running and providing service to a population of subscribers. The actual number of subscribers serviced is a function of the network's designed capacity. Other users experience degraded service, such as blocked calls or blocked short messages. SS7 implements congestion handling through buffer monitoring and cooperation between the SS7 transport and user parts. MTP2 employs a series of message buffers for links and for message passing between different SS7 layers.

Operators can define buffer settings that trigger SS7 messages. These settings usually have three values:

- **Congestion Onset Threshold** – When buffer occupancy passes the onset threshold, the SS7 transport part (typically MTP2) sends a congestion onset message to the user part (the application using the SS7 network). When the user part receives this message, it is expected to throttle back or discard messages.
- **Congestion Abatement Threshold** – When buffer occupancy drops below the abatement threshold, the SS7 transport part sends a congestion abatement message to the user part. When the user part receives this message, it stops throttling traffic from the application.
- **Discard Threshold** – When buffer occupancy passes the onset threshold, the SS7 transport part discards messages.

The above approach is consistent with SS7's congestion handling philosophy: stem the traffic flow as close to the source as possible. While the SS7 transport part detects congestion, it does not handle it directly. Instead, it asks the user part to invoke its congestion handling routines as it is closest to the application and can implement throttling algorithms suited for that application.

Unfortunately, this flexibility means congestion handling is implementation-independent [21]. All SS7 networks provide the capability to detect congestion, but the user part plays the largest role in handling congestion. Research has been done to determine optimum SS7 settings for wireless networks [22]. However, these efforts usually focus on optimizing performance for the general population, rather than specific NS/EP needs such as wireless priority service.

### 3.2.2.5 Short Message Service Controller (SMSC)

While the SMSC is not covered in the scope of GSM, CDMA, or other wireless specifications; its behavior is part of the TIB. Implementation of congestion handling routines is in SMSCs and is vendor-dependent. Typically there are three approaches that vendors take to handling congestion:

- **Application Throttling** – Application Throttling applies to applications that use the SMSC to send messages depending on the level of service given to each application. Throttling is accomplished through a flow control mechanism. Every message submitted to the SMSC has an acknowledgement and the SMSC will pause before sending the acknowledgement. This delay is proportional to the throughput it wants to provide to the application. Messages sent during the pause are dropped by the SMSC. For an application to work properly, it must wait for its proper acknowledgement from the SMSC.
- **Buffers** – The second technique for handling congestion is through a series of buffers. These buffers hold messages in user unique partitions. If a handset is off, the message is stored in the user's message buffer. The buffer also will fill up if messages are arriving faster than the network can process them. When the buffer is full, the SMSC typically stops accepting messages for that user until his buffer length decreases. Buffers also are assigned to individual applications, such that an application sending messages into the SMSC has a limit on the number it can send without them being delivered.

Sample sizes for buffers can be on the order of hundreds of messages for individual users whereas an application might have a total of 500 messages. If either is full, the SMSC won't accept a message sent from an application to a user.

- **Limits on Total Throughput** – The third method of controlling congestion is based on throughput. Typically when SMSCs are sold, they are licensed on a message per second basis. When an operator buys an SMSC, the equipment may be capable of delivering more messages than the software license allows, however the software won't deliver more messages per second than the operator has paid for. Therefore the traffic is throttled by the software license. This provides a clear upper bound across all traffic for an SMSC, so it is not possible for traffic to exceed this licensed limit on the SMSC regardless of user type.

### 3.3 Vulnerability Assessment

A number of companies provide advanced services and are allowed to connect directly to the SS7 network. These bring additional vulnerabilities and security liabilities on a case-by-case basis. These systems can possess a public Internet interface, which offers potential hackers an additional avenue of attack.

As with any system connected to the public Internet, it should be assumed that at some time, the SMS system would be the target of a malicious attack. Whether the attack is caused by a group of hackers seeking simply to disrupt the system or is sponsored by a group of individuals attempting to spy on and mislead competitors, the SMS system must be protected from outside attacks. As with any complex system composed of multiple parts, there are numerous

vulnerabilities that should be examined thoroughly in an effort to predict upcoming attacks, and defend against them as they occur.

### **3.3.1 Denial of Service (DOS) Attack**

All machines with public IP addresses attached to the Internet are susceptible to a Denial of Service (DOS) attack. A DOS attack is effectively an attempt to disable or render ineffective a network-based service by deliberately bogging it down with excessive network traffic. DOS attacks gained notoriety in early 2000 when a group of hackers managed to shut down Yahoo.com for approximately two and a half hours. In that instance, the hackers used a distributed DOS attack, a variation of the attack in which an attacker infects up to millions of individual computers with a virus and then coordinates a strike in which each virus-infected machine begins to rapidly open and close connections to a specific server at the same time. In the case of Yahoo, millions of infected computers around the world simultaneously began to assault Yahoo's web servers with 1 Gigabit of data per second. Distributed DOS attacks are among the most difficult to defend against because any personnel computer in the world could be carrying the virus. If an operation depends on most computers being able to connect to it (such as most web and mail servers), an administrator cannot use a firewall to filter out previous attacking computers without alienating their user base. Additionally, it is almost impossible to find the attacker who originally designed and distributed the virus.

The majority of Denial of Service attacks do not result in the halting of the target service. These attacks, however, still waste a large amount of bandwidth and bog down the efficiency of the attacked machines. As a result, even the small attacks that occur hourly around the world do significant damage to the efficiency of internet-based services.

Most SMSCs have a slight advantage in defending against DOS attacks in that the SMSC itself is often shielded from the public behind a firewall. Although Denial of Service attacks against a firewall directly increase network latency and waste bandwidth, firewalls can be configured to ignore and not respond to errant packets of data. Systems outside the SMSC, such as an email server and other applications that connect to the SMSC from the public Internet are susceptible to DOS attacks, and the individual features those systems provide are capable of being disrupted. Under this layout, while a DOS attack can disrupt the internet-based features of the SMS system, the SS7 features of the SMSC such as message storage remain intact.

### **3.3.2 Service Interruption Attack**

A Service Interruption attack is similar to a Denial of Service attack, but varies in that it uses a smaller number of oddly formatted messages to attempt to take advantage of bugs in the target to disrupt a service. These attacks are more refined than a DOS attack and often depend on in-depth research or insider information to determine exactly how to format a message to produce a chaotic result in the target application.

The majority of these attacks are executed by individuals who were involved in the design and production of the system they attack and design the attack based off of vulnerabilities they found in the code. On some occasions, the individual intentionally placed such design faults.

These forms of attacks also occur frequently during the testing stages of open source software, although the large number of contributors and reviewers in the open source community help eliminate the vast majority of these deficiencies before the software reaches a release version. Commercial software is often rushed and poorly tested, and the number of persons participating in code reviews is often much smaller statistically leading to more potential holes. The fact that the code is hidden from the public indicates that such vulnerabilities can remain undetected for the lifecycle of the product.

These vulnerabilities are sometimes found by accident, as was the case with the infamous “WinNuke” vulnerability of 1997. The WinNuke vulnerability was a bug within the TCP/IP protocol stack of the initial release of Windows 95 which, when receiving a specifically formatted TCP header, would cause the entire system to crash. While the bug was fixed relatively quickly, any PC connected to the Internet and not protected by a firewall would crash within minutes of connecting due to the large number of hackers flooding the Internet with these fatal packets. The earliest source to write about the WinNuke bug found the faulty TCP header accidentally while writing drivers for a router.

Not all Service Interruption attacks are designed to crash the target machine. Often they are simply data inputs that the attacker knows will cause any form of chaotic and unpredictable result on the target. There are countless bug reports of various systems that, after receiving faulty input either deliberately or accidentally, needed to be restarted or reconfigured.

Service Interruption attacks are difficult to predict from a macro level. They depend on the individual implementation of the system. A well-designed and implemented SMSC could be built as to be virtually immune from this form of attack. It is important to note that SMPP and the other protocol by which the SMSC communicates are simple and closed enough that holes do not exist in the protocol itself which could cause a system to be interrupted.

### **3.3.3 Service Hijacking Attack**

The most notable form of information attack is Service Hijacking, called “cracking” in the hacker community. Service Hijacking occurs when an individual gains control of a process on a machine to which they should not have access, or in the worst-case scenario, when the individual gains control of the entire machine. The most dangerous aspect of this form of attack is that once an attacker gains control, they can deliberately modify data to achieve a goal of their own, most often counter to the goals of the service’s operator.

The dangers of having a service hijacked depends on the target services functions. In the case of systems like the SMS network, which is used to push private messages to cell phones, the principle concerns are that an unauthorized person will be in a position to read and/or modify messages as they pass through the system, and that the person might attempt to read or modify data on the connected SS7 network. An infiltrator can also modify the software on the target machine to allow easy access at a later time, or to transmit a certain type of data to an external listener as it is encountered. Infiltrators can also sabotage the system itself, making it run less efficient, or destroying the capabilities of the compromised machine to operate at all.

Two of the most common methods of hijacking an application are buffer overflow attacks and password compromise attacks

### **3.3.3.1 Buffer Overflow Attack**

The Buffer Overflow attack is a vulnerability caused by a common bug in many applications written in C, C++, or assembly language. This flaw is called an unchecked input bound error, and it occurs when the developers of a piece of software do not ensure that there is room in the memory buffer used to store data read from the network for an incoming message. If no such check is done, in certain conditions an attacker can send a message significantly larger than the buffer in question and fill the excess data with malevolent machine code. This excess data can often overwrite a portion of the existing program, in effect allowing the attacker to reprogram it. The next time that region of machine code is accessed and executed, the instructions of the attacker will occur in place of the intended program, and if the attack is planned correctly, the attacker will gain control of the program. The bugs that cause this vulnerability are frighteningly common and have been found in a large variety of services such as Microsoft's SQL Server and the Apache Web Server.

### **3.3.3.2 Password Compromise Attack**

The most common and most difficult to detect attack is password compromise attacks. In this form of attack, the attacker typically accesses the target machine using normal channels and gains access in a seemingly inconspicuous manner. The majority of password compromise attacks are the result of "social hacking", in which the attacker gains a user's account information by obtaining their trust. Passwords can also be obtained via "brute force", or the process of trial and error. Given enough time, a system can algorithmically generate the correct account information.

In regards to the SMSC, like any other system, this form of attack is generally the most dangerous. As the SMSC is normally shielded from the public Internet and only allows connections from a pre-specified list of machines, the majority of the dangers to the SMSC are alleviated since an attacker must directly connect to the SMSC in order to hijack it. However, a determined attacker can employ a "piggyback attack", in which the attacker takes control of a machine on the public Internet that is also connected to the SMSC, such as a mail server, and then uses that machine to attempt to hijack the SMSC.

### **3.3.3.3 Snooping**

Another method obtaining passwords is by a process called "snooping". Snooping is effectively spying in any manner on a data stream and successfully extracting information. As individual users begin to exchange more personal data via SMS, the need for data privacy becomes increasingly important. Snooping is normally accomplished in one of three ways.

- The first method is via a Service Hijacked system on the existing network. In this case, as described above, an attacker has acquired control over a system with access to a confidential data stream. This method also includes "insiders", or administrators or operators of the system in question who abuse the power given to them.

More disconcerting is that banks and other financial institutions are beginning to implement security systems using SMS systems. Additionally, corporations are beginning to implement One Time Password solutions as a method of securing network resources. A One Time Password over SMS is a system in which a user attempts to log into a specific resource by entering their username; instead of entering a memorized password, however, the system forwards a unique password to their cell phone that is only valid for a short period of time. One Time Passwords are a proven security feature that virtually eliminates social password compromises, but SMS implementations of the system are fully compromised if the user's cell phone is stolen, or if an alien machine is snooping for data on the SMS network. As mentioned above, snooping a network can allow one to record the login information of other valid users, giving the attacker an opportunity to continue their operations at a later time.

- Second, snooping is often also accomplished by external systems, via Link Hijacking, along the data path. In a TCP/IP network environment, data is routed from its origin to its destination by being transmitted across any number of subnets. As the data is transmitted over a link, it is received by every other device connected to that link, including the next machine in the routing sequence for which the data is intended. That machine receives the data, identifies from the IP header that the data is intended for it, and then forwards the data into a separate subnet destined for another routing entity. This process continues until it arrives at the intended destination device. While network data was received by every other device on every wire along the routing path, these machines typically ignore the data after determining from the IP header that they were not intended recipients. These machines can be programmed, however, to record this information. It is a common tactic for hackers to attempt to Service Hijack a device along a common route to a service or on a nearby subnet to the service. Using these hijacked machines, the attackers would be capable of viewing and recording data as it passed into and out of the SMSC. In this case, although the SMSC is the actual target of the attack, the intrusions into the network are done by connecting to other machines besides the SMSC, such as the WAP or email gateway service machines. Encrypted communications alleviates the majority of concern regarding spoofing, but dedicated attackers can also attempt to break the encryption scheme by stealing the keys if they are not well protected.
- And thirdly, Connection Hijacking is a process in which a snooping machine begins transmitting data to a listening machine in an attempt to “convince” that machine that the data is actually from an existing valid connection from a third computer. Connection Hijacking is especially dangerous because it allows an attacker to wait until an authorized connection has been formed and login account information exchanged before it steals the connection, giving it all of the abilities of the previous connection.

Fortunately, Connection Hijacking is a complex and often computationally heavy operation and can be difficult to perform in real-time. First, the attacking machine must be in a position to snoop a connection (it must be connected to the network along the route that a stream of data from a source machine and the target would follow). Secondly, if the connection is encrypted, the hijacking machine must possess the encryption keys, or it cannot encode and decode data in a timely manner. The third criterion is the most difficult: TCP hijacking is performed by the attacking machine listening to the data sequence and finding a pattern in the TCP identifier field for each packet passing through the stream. If a pattern is found, the attacker must predict a future value of this field and wait until just before the target machine would send a packet with an identical value. At that moment, the machine must begin to

transmit data using the original source machine's IP and MAC address in the TCP/IP header and starting with the TCP sequence number predicted. If this packet arrives at the targeted machine in the window between when the packet with the previous sequence number arrived and when the spoofed packet arrives, then the server will respond to the packet from the attacking machine and ignore the data from the original source. The attacker must then disrupt the original sending machine from sending more data, as conflicting data with identical TCP/IP header information could cause intelligent protocol stacks on the targeted machine to recognize that a problem had occurred. If an attacker is capable of performing all of these tasks, and is aware of the protocol used to communicate with the target machine, the connection can be hijacked successfully.

### **3.3.4 Spoofing**

All of the above attacks almost universally apply to any public machine on the Internet. The nature of the SMS system opens a few unique security issues, the most critical of which is spoofing. Spoofing is the act of sending a valid data message, but obfuscating the sender so as to hide the true origin of the message and, in some cases, pretend the message originated from a different source. The internet-side SMS features such as web page based messages and email-based messages have no system of authentication, allowing any person to send messages that can appear to be any other Application Originating Message. With no form of authentication, an end user is forced to determine whether or not to trust the source of a message they receive based solely on their own individual judgment without technical aid. In the case where certain trusted users are granted direct access to a provider's SS7 network, a hacker could send a malicious control message to the network under the guise of the trusted user.

### **3.3.5 Non-Internet Based Methods of Attack**

This section addresses three non-Internet based methods of attacking a network. The goal of these attacks is to deny service either by causing congestion in finite resources or by rendering the network inoperable. The techniques described include causing congestion through:

- Voice calls,
- Congestion through short messages, and
- RF Jamming

#### **3.3.5.1 Voice Calls Congestion**

The first method of attacking is to cause congestion in a network by placing a large number of voice calls either from PSN, or from cell phones located in the area where the congestion is desired. There are commercial, off-the-shelf (COTS) products available today that can automate this process for both cellular and PSN phones. This technique is fairly simple and inexpensive to implement. A disadvantage to this technique is that the source can be traced either to the cellular subscriber or to the connection into the PSN. This type of congestion, while not the intent of the attack on the Pentagon, was experienced in the Washington, DC area during the hours after the events of 11 September 2001.

### 3.3.5.2 Short Messages Congestion

The second method of attack involves generating a large number of short messages using cell phones to cause congestion of the SDCCH capacity in a particular area. There are COTS products available that can automate this process. One advantage of this method is that it is simple to implement. A disadvantage is that the phones used to generate the messages can be traced.

### 3.3.5.3 RF Jamming

The third method of attack is an RF jamming. Jamming equipment is placed in the area where the service is to be denied. This equipment transmits a noise signal that causes interference in the base stations in the wireless networks and prevents them from clearly receiving transmissions from cell phones. It is well known that IS-136, GSM, and iDEN systems are easy to jam. CDMA-based technologies are based on spread spectrum principles and therefore are more difficult to jam; their anti-jamming properties are well documented.

One advantage of this technique is that it would affect all wireless traffic (SMS or voice) within a system. Another advantage is that it is simple to implement; all pieces needed for a solution can be obtained commercially. A portable solution designed to fit in a backpack could be very effective. Another advantage is that mobile jammers are not easily traceable.

The main disadvantage of this approach is that it is very location and technology sensitive. While it would be straight forward to jam one or more networks in a specific area, it would be much more difficult to jam all of the networks across a large geographic area.

## 3.4 SMS-Related Developments

### 3.4.1 Cell Broadcast

Cell Broadcast Service (CBS) is a relatively new feature designed to transmit identical text messages to cellular phones covered by a specific cell site. Users can 'tune in' to different cell broadcasts and receive all information broadcast to that cell. Two key benefits of this functionality are: (1) the ability to send a single message to a large population in a very efficient manner; (2) the ability to direct messages to a specific geographical area. Table 3-1 summarizes key differences between SMS and CBS.

**Table 3-1. SMS and CBS Key Differentiators**

| Short Message Service   | Cell Broadcast Service   |
|---|--|
| <b>Location Independent:</b> a message reaches the recipient, regardless of where they are. | <b>Location Dependent:</b> a message is sent only to a specified location.   |
| <b>Unicast:</b> a message is sent to one person at a time.                                  | <b>Broadcast:</b> a message is received by anyone "tuned in" to the channel. |
| <b>Use:</b> person-to-person messaging.   | <b>Use:</b> emergency alerts, traffic data, etc.                             |



Because the same questions surrounding SMS and NS/EP may also apply to CBS, a brief investigation of CBS is included. While CBS was designed to meet requirements that conventional, point-to-point SMS could not, there are some similarities between the two technologies. Both are text-messaging technologies and require a message center node attached to the wireless network and use the fixed network to transmit messages between the message center and the wireless interface.

#### 3.4.1.1 CBS Usage

Today, CBS is commonly used to provide services such as:

- Weather and traffic information
- Stock quotes
- Horoscopes
- Headline news and sports
- Airport information

For some applications (weather, traffic, airport), the location-based qualities of CBS are employed. For other services (stock quotes, horoscopes, headlines news and sports), broadcast efficiencies make CBS the only viable technology to provide information to a large population [23]. CBS could also be used in support of various emergency services [24], such as:

- **Emergency Information Services** — During local emergencies, a public authority might broadcast a message to very specific areas. This message could:
  - Alert users in the path of severe weather events such as tornadoes
  - Alert users near environmental hazards such as chemical spills
  - Alert users to location-specific terrorism information
- **Emergency Workers** — During local emergencies, the appropriate authority could use Cell Broadcast to reach on-call employees in a specific geographic area.

#### 3.4.1.2 How CBS Works

Most wireless equipment vendors provide Cell Broadcast Centers (CBCs) separately from their SMSC offering. The interface between the CBC and BSC is not specified, but is left as an implementation detail. Guidelines suggest multiple implementations, one of which uses SS7 [25].

Cell broadcast messages tend to be shorter than SMS messages. In GSM's implementation, cell broadcast messages are 82 bytes and, depending on character coding, may hold up to 93 characters. Multiple messages can be concatenated to create longer messages.

Cell broadcast messages are resent periodically. The broadcast period can be specified for each message and is usually set according to message content (e.g., traffic updates might be broadcast more frequently than weather updates). On GSM networks, the rebroadcast period may vary

between 1.88 seconds and 32 minutes; a message can be rebroadcast indefinitely, or between 1 and 65535 times. Messages contain a message identifier and sequence number, allowing a phone to determine whether messages are new or re-broadcasts of messages received previously.

Cell broadcast messages are placed into categories (classes). A user can decide what types of broadcasts to receive by instructing his phone to “tune in” to only those categories. For example, a user might want to receive traffic information and local sports, but not weather. In this case, they would direct their handset to receive broadcasts from two classes but ignore a third class. Finally, cell broadcast messages can be transmitted to cells, lists of cells, location areas, or the entire network [26].

### 3.4.2 Enhanced and Multimedia Messaging

Over time, the nature and form of cellular communication is getting less textual and more visual. GSM messaging has evolved beyond text by taking a development path from SMS, to Smart Messaging (introduced by Nokia), to Enhanced Messaging Service (EMS) and finally to Multimedia Messaging Service (MMS). Table 3-2 provides a summary comparison of the GSM messaging technologies.

**Table 3-2. Comparison of GSM Messaging Technologies [27]**

| <i>Type</i>          | <i>Characteristics</i>   | <i>Content Formatting</i> | <i>Applications</i>  | <i>Support</i>                              | <i>Timeframe</i> |
|----------------------|--|---------------------------|--|---|------------------|
| Text Messaging       | 100-200 characters   | Yes                       | Simple person to person messaging                                      | All Phones                                  | 1990s onwards    |
| Smart Messaging      | Simple rudimentary images and ringtones  | Yes                       | Simple person to person messaging with a visual feel                   | EMS standards expected to be widely adopted | 2001 onwards     |
| Enhanced Messaging   | Text messages plus simple media formats e.g. sound, animation, picture, text formatting enhancements | Yes                       | Simple person to person messaging with a visual feel                   | EMS standards expected to be widely adopted | 2001 onwards     |
| Multimedia Messaging | Messages in multiple rich media formats e.g. video, audio plus text                                  | Sometimes                 | Person or server to person messaging with rich image and video content | MMS standards becoming widely adopted       | 2002 onwards     |

#### 3.4.2.1 Enhanced Messaging Services (EMS)

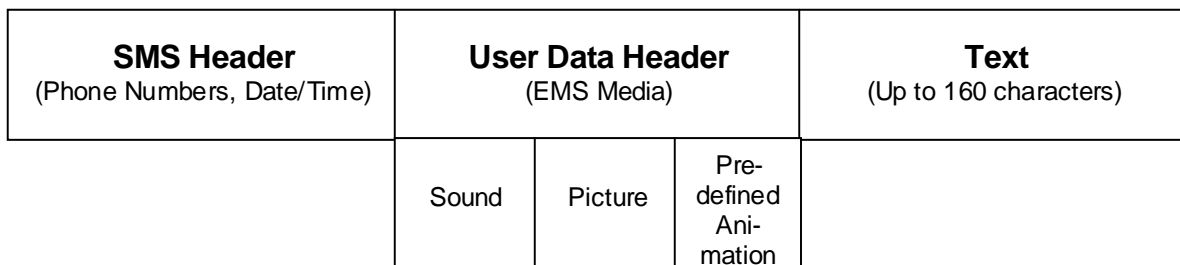
The Enhanced Messaging Service is a standard developed by the Third Generation Partnership Project (3GPP) [28] to embrace and extend the ability to send ringtones and operator logos and other simple visual messages to EMS-capable handsets. EMS enables users to send and receive a

combination of simple media such as melodies, pictures, sounds, animations, modified text, and standard text as an integrated message for display on an EMS compliant handset [28].

Enhanced Messaging Service is the natural progression from SMS's 160-character text messages. EMS is built using the existing SMS infrastructure. It began as a cross-industry collaboration between the handset manufactures Ericsson, Motorola, Siemens, and Alcatel. By leveraging existing SMS infrastructure, it keeps investments to a minimum for operators and provides a familiar user interface and compatibility with existing phones and with other manufacturers [28].

The Enhanced Messaging Service has been standardized by extending the use of User Data Header (UDH) in the GSM SMS standard. The UDH makes it possible to include binary data in a normal short message prior to the text itself. EMS is an enhancement to SMS but is very similar to SMS in terms of using the store and forward SMS Centers, the signaling channel to realize EMS. EMS has little or no impact on today's SMS Centers. The introduction of EMS should be totally transparent to SMS Centers since they already support the User Data Header. This is a key advantage to EMS, since network operators do not need to make additional investments to SMS Centers or network infrastructure assuming their networks already support binary 8 bit messaging and unless EMS message volumes mean investment in new SMS Center capacity.

In an EMS message, all the media is stored in the User Data Header (UDH), while the text is stored in the normal text area of the Short Message. This is shown in Figure 3-6.



**Figure 3-6: Short Message Schema [29]**

As pictures and sound can be included in the message it is easy for the users to generate a single EMS message containing a large number of SMS messages. Hence, because messages can be concatenated, there is no 160-character limit. However, the current ratio between EMS messages and SMS messages weighs heavily toward SMSs, and the addition of EMS messages to the traffic load across SMS centers and networks is not seen as a major impact.

### 3.4.2.2 Multimedia Messaging Service

Multimedia Messaging is a further migration from SMS. It supports media containing Joint Photographic Experts Group (JPEG), Graphics Interchange Format (GIF), text, AMR voice and

other formats. MMS also supports email addressing and interoperability. Like SMS, MMS is an open industry standard and is carrier independent. [30]

Specific media formats and support that are supported include the following:

- Text – unlimited text that can be formatted.
- Graphics – support for graphs, tables, charts, diagrams and layouts.
- GIF – support for animated GIF's.
- Audio – support for music and speech; and streaming sound.
- Images – the ability to send images and snapshots from an attached or built-in digital camera.
- Video – ultimately, the ability to send video, clips or streaming (over a full network).

The steps for sending a message across a Multimedia Messaging Service Center (MMSC) are as follows:

- Sender sends a message to the MMSC.
- MMSC sends a confirmation of receipt and the sender gets a 'message sent' notice.
- MMSC sends the receiver a notification that a new message is waiting (via SMS).
- Receiver downloads the message immediately or later (via WAP).
- Once the receiver has successfully downloaded the message, the sender gets a 'Message Delivered' message.

For MMS to be implemented the network operators have to upgrade their infrastructure and devices supporting MMS must be made available by the handset vendors. Unlike SMS, which does not require voice or data channels, MMS uses data (in GSM it is typically GPRS) and signaling channels. Both SMS and MMS are store and forward systems and are not real time. However, unlike SMS, MMS can use user profiles to determine when content should be delivered based on availability or time. MMS can also undertake format conversion based on terminal characteristics and user profile. MMS is based on SMIL (Synchronized Multimedia Integration Language), where the presentation information is coded into the presentation file. This allows the multimedia content to be presented in a specific order at a predetermined interval.

The WAP Wireless Service Protocol (WSP) is used as a transport mechanism in MMS (although use of the WAP browser is not required). By using WAP as a transport, any bearer with WAP capabilities can be used, which makes the MMS bearer independent. Consequently, MMS is not limited to only GSM.

There are some clear differences between the MMS environment as compared with SMS. In the SMS environment, storage of messages is not normally an issue. This is due to the size (length) of the SMS message (160-characters maximum). The storage requirement only becomes an issue

when the recipient is not available and when the collective number of recipients needing storage because of their unavailability is significantly large. Because of the dynamic nature of the store and forward feature as a function of availability of the recipient, store and forward issues are rarely seen.

In contrast, MMS messages are normally larger. They may also be stored in the recipient's MMSC for longer because they may not be immediately downloaded and because users often want to keep the stored message for longer periods of time or even permanently. Consequently, operators and vendors supporting MMS must initially plan for larger and potentially expanding storage needs. Ultimately, MMS creates a highly interoperable environment where cell phones become an equal player with other sophisticated PDAs and the Internet.

### **3.4.2.3 EMS and MMS Implications to NS/EP**

The fact that SMS has evolved to EMS and MMS might suggest network traffic loading vulnerabilities, particularly in time of peak or surge usages such as times of crises – terrorist attacks, national emergencies, public safety events, and natural disasters.

From a traffic loading perspective, EMS represents the concatenation of SMS messages. These longer SMS messages have a slightly higher impact on a network than standard SMS traffic. However, EMS represents a minor portion of total traffic today and should not increase noticeably in the future. Consequently, there seems little reason to explore the increases in network vulnerabilities due to EMS usage.

MMS behaves more like voice traffic than SMS traffic. MMS data is carried over the traffic (normally packet data) channels. While some MMS traffic is growing on the 2.5G (GPRS) networks, its real growth will occur with the roll out of 3G networks specifically designed to handle the high data rates needed to support large volumes of MMS traffic.

In conclusion, the low take-up of EMS and different traffic characteristics of MMS are not expected to affect NS/EP functions significantly more than SMS could. However, EMS and MMS could be of great use during period of crises when vital data beyond text messaging is needed.

### **3.4.3 Current and Future SMS Services**

There is little doubt about the success of SMS across the globe among individual users. Within the last two years, early inroads have been made in various enterprise and government markets and are now beginning to show real growth trends.

In Europe alone, messaging volumes are well over \$10 billion a month, with Asia adding huge volumes for a combined revenue realization of over \$25 billion worldwide at the end of 2002. SMS messaging in the United States has lagged far behind Europe and Asia largely due to the lack of harmonized technology standards enabling interconnectivity and exchange of messages among consumers utilizing competing carrier networks. Moreover, well established wireline and Internet infrastructure, low wireline rates, and free Internet usage have worked to further depress wireless messaging. [31]

However, with the strides that have been made in bridging the interoperability chasm, specifically the enabling of US inter-carrier messaging in April 2002, the emergence of SMS service providers and the introduction by the SMS Forum of Common Short Codes (CSC), SMS messaging is beginning to show real growth in the America. Additionally, handset manufacturers began to embed messaging capabilities in their new products, and today 99% of cell phones sold in the US market include text-messaging functionality. [32]

From a business perspective, carriers are now including messaging capabilities in many of their service packages, instead of requiring separate sign-up. These initiatives have resulted in an immediate uptake in message volume growth, with some carriers reporting message traffic increases of up to 400% [33, 34, 35].

The Yankee Group has projected that the penetration rate for active SMS users in the U.S. will grow from 9% in 2002 through 30% in 2004 to 47% by 2006.

### 3.4.3.1 Commercial Sector Services

Commercial usage of messaging (every day users) continues to dominate the generation of SMS messages and will for years to come.

- **Person-to-Person 2-way Text Messaging** – This continues to lead in generating the bulk of SMS traffic in the US. Simple person-to-person messaging is used to say “hello”, to prompt someone for something, to arrange a meeting, etc. The response can be equally short, and often prompts a further response, and so forth. Additionally, vendors have made great strides in implementing predictive text features into their handsets. Predictive text input algorithms such as T9 from Tegic significantly reduces the number of key strokes required to input a message. These kinds of features are accelerating the use of SMS messaging. [36]

While it began as pure SMS messaging, Instant Messaging (IM) is beginning to add to the growth as well. By the end of 2002, Cingular Wireless had reported increases of several hundred percent, and AT&T and Verizon have also shown sizeable increases. The services offered by these and other carriers include 2-way text messaging, SMS to email, ringtones, games, and sweepstakes. [37]

- **Information Services** – SMS messaging is used to deliver a wide range of information to cell phone users, including stock quotes, sport scores, weather, travel and flight information, new headlines, lottery results, jokes, and horoscopes. In these commercial applications, information is normally configured in a push-based format triggered by time, events, or conditions (e.g., the football score right after a touch down).
- **Email and Internet Email Alerts** – One of the most common uses of SMS is for notifying cell phone users that they have new voice or fax mail messages waiting. Whenever a new message is dispatched into the mailbox, an alert by SMS informs the user. Email can be sent to a handset when the customer’s cell phone number becomes a part of the email address. Emails sent to that address are forwarded as a short message to the wireless phone. While attachments are not included, the user is alerted of the need to go on line and access the

attachment. This application is and will remain one of the largest generators of short messages.

- **Ringtones** – The availability of polyphonic cell phones and the ability to download a ringtone is an SMS-based application that is growing in the US. Users may download ringtones from their carriers, or from Internet site. Ringtone composites are also popular because they allow cell phone users to compose their own unique ringtones and download them to their phones.

### 3.4.3.2 Enterprise Services and Applications

Enterprise services and applications have begun to be implemented in the last year, and, as SMS evolves to EMS and MMS, the enterprise applications will reflect significant growth. The following are areas where the greatest impact on SMS message generation is or will most likely be felt. [38]

- **Travel** – time-critical business information and applications about travel to employees, customers, and partners across a variety of devices. Services are beginning to be offered in customer service, marketing and promotions, and booking and reservations.
- **Field Service Support** – time-critical business information and applications to employees, customers, and partners in the office equipment, utilities, telecommunications, medical, industrial and pharmaceutical markets.
- **Finance** – services covering market information, wireless promotions (marketing campaigns, sales and advertising), transactions and enterprise productivity applications
- **Media** – Interactive television (e.g., using SMS to vote for American Idol contestants), gaming, online content markets and other media for broadcasting, content delivery, events and ticketing and transactions

### 3.4.3.3 Services Relevant to National Security and Emergency Preparedness

The most recent trends in wireless messaging are the emergence of applications in the public sector. The leading application in this area is the use of alert and notification broadcasts in a variety of scenarios, but the ability to send and receive Instant Messages and conduct chat are also applications that will emerge and demonstrate their value.

- **Alert and Notification Broadcasts** – These are push messaging applications that provide instant alerts to wireless devices that could be located in one region or widely dispersed. They support both government and non government personnel to include but not limit to:

Government or quasi-government personnel at local, state, regional, and national levels:

- Law enforcement
- Federal agencies
- Intelligence community
- Homeland security
- Port authorities (national and international)

- Emergency response agencies (e.g., FEMA, Red Cross)
- First responders (e.g., emergency medical services, firefighters)
- Government disaster teams
- Embassy personnel
- Liaison officers between agencies

Non government:

- Volunteers to first responders
  - Volunteers to emergency and disaster teams (e.g., individual medical experts, amateur radio teams, civil air patrol)
  - Citizens
- **Instant Messages and Chat** – This application can provide time-critical communications across internal and external management, analyst, and operational teams around both normal and crises situations where other, more efficient forms of communication are not available. Once an event has been identified and operation is underway, instant messaging and the creating of specific chat rooms for the event could bring increased information exchanges and still support time-sensitive decisions. The value of these capabilities will be greatly enhanced with the implementation of secure networks and devices.

There are other uses within the public sector for SMS. The use of voting by cell phone is now in trial in certain parts of Europe (e.g., England). Another use is in polling – collecting the data for public surveys concerning key political issues. [39]

Ultimately, many new applications needed to support national security and emergency preparedness will emerge. These applications may bring the strength of instant communications with selected groups of people or agencies – as determined by the message sender, coverage ubiquity, collaboration, interoperability and eventual selective security in communications as needed.



## 4. Security

As with any form of data communication, security is becoming an increasingly important concern with SMS. This section reviews the protective measures SMS employs for sender authentication and privacy.

As a rule, an SMS message's data privacy relies on the following:

- Security of the destination device's network
- Message source

SMS inherits and utilizes security mechanisms in the carrier's network, rather than relying on special security features. Non-standard, proprietary protections to SMS messages are generally not available. It is unlikely that new authentication or encryption standards for SMS will appear in the future, as current security is considered "good enough" for casual use, since new algorithms would require network and handset changes (a costly proposition) and interoperability problems would have to be considered.

From the point of view of message privacy, an SMS message passes through up to three significant phases which must be examined: within the carrier's SS7 network, over the air, and potentially over the public Internet. All messages pass through the first two states, while messages that originate from applications or users on the Internet pass through the public Internet. Therefore, we analyze the following two types of SMS messages:

1. Messages that originate from users or applications outside the carrier's network on the public internet (Application Originated Messages)
2. Messages that originate from another cell phone (i.e., mobile originated messages).

### 4.1 Application Originated Messages

Application Originated Messages originate on a connection outside the wireless carrier's network. These messages are transmitted using a plain text protocol such as SMPP, UCP, or CIMD to a public IP address. This IP address could represent the SMSC itself, or another machine acting as proxy between the Internet and the SMSC. This communication is not encrypted and is not considered secure. As with any plain text TCP/IP connection, these communications can be read by a system on the routing path between the message sender and SMSC. They can also be modified through a process called "hijacking".

Most SMSCs are protected by a white list. White lists are lists of IP or MAC addresses of machines from which a server will accept connections; the server denies all others. Some SMSCs also require a username/password combination before allowing messages to be sent from that specific connection. However, some SMSCs have no security features at all and will accept connections from any machine on the Internet. While these security features help prevent Spam and large volumes of anonymous messages from being injected into the wireless network, they do not protect the privacy of the messages being passed.

In addition to direct SMSC connections, most wireless carriers allow sending an SMS via email sent a user's cell phone address. These systems are implemented using an email server on the public Internet connected to an SMSC. After an email is received, it is pushed to the SMSC via an SMPP (or other) connection. From this point on, the SMS follows the path of other Application Originated Messages. This approach is one of the most popular methods to send SMS messages to cell phones. These short messages suffer the same drawbacks as Internet email:

- Messages are not private.
- By modifying the sender's address, a user can masquerade as any sender.

As a result of the privacy concerns, some SMSCs allow incoming connections to be "Secure Shell (SSH) tunneled" into the SMSC. SSH tunneling is a method of encrypting data communications by routing them through an SSH application on the receiver's machine. Rather than connecting directly to an SMSC on the conventional SMPP, UCP, or CIMD port, the sending application connects to a separate port listening for SSH connections on the SMSC. The sender then provides SSH with the SMPP, UCP, or CIMD port number, and the SSH receiving program receives the data, decrypts it, and then passes it to the destination port using shared memory buffers.

It is also possible to use a two-machine approach in which the SSH server and SMSC are two different servers connected by a private connection. In this case, after the data arrives at the SSH server, it is transmitted in a decrypted form on the private connection to the SMSC. In both cases, the data is protected and remains private. Depending on implementation SSH may use several encryption algorithms. All are strong, shared key algorithms. Tunneling provides an efficient and highly effective method of protecting the data contained within messages. Since tunneling does not require any modification of existing SMSC software and works with any TCP/IP-based protocol, it provides effective security while maintaining interoperability and backwards-compatibility.

In regards to the email solutions mentioned earlier, if SSH tunneling were used to send the initial email to the mail server and SSH tunneling were used to bridge the mail server and the SMSC, the transit of the message over the public Internet could be considered private. Additionally, SSH tunneling can support username/password fields and public key systems as authentication methods, although key management systems would require vendors to deploy custom, non-standard software.

## **4.2 Mobile Originated Messages**

Messages sent from one cell phone to another cell phone never leave the carrier's private network and have security advantages compared Application Originated Messages. However, at some point, all messages are transmitted over the air, which introduces additional security issues. For transmission over the air, there is no significant difference between sending and receiving security.

After an SMS message has arrived at the SMSC, it must be routed to the appropriate switch and then to the appropriate base station for transmission to the cell phone. The SMS is transmitted from the SMSC to the MSC using SS7 messages. The SS7 network does not encrypt data before transfer, but the entire network is separated from the Internet and protected against access. All carriers control access to their SS7 networks very heavily, and even restrict who may view data on the networks. Intrusion into the networks requires direct access to the service provider's equipment, or the installation of a listening device at some point along the network's wired connection. This makes SS7 secure enough for most users, although it does not meet FIPS requirement for sensitive information.

After a message has been successfully routed to the base station for transmission to the mobile, it is broadcast over the air. At this point, the message leaves the fixed network and must be protected by other means. In the United States, there are four major types of digital wireless networks – CDMA, iDEN, GSM, and TDMA. All possess strong authentication and routing mechanisms to ensure messages are not accidentally passed to the incorrect mobile. In addition, each has some level of scrambling or encryption for message privacy.

CDMA is the technology used by Sprint PCS and Verizon Wireless. Although the standard for CDMA supports encryption, most networks do not activate this feature. Rather, the spread spectrum nature of CDMA, combined with long spreading codes, scramble the communications and make it difficult for an eavesdropper to snoop on voice calls or SMS data. However, these scrambling techniques are not as effective as true encryption. The time and effort involved to decipher encrypted data provide a protective barrier. However, deciphering scrambled data is significantly easier, and as such, only acts as a deterrent. If in the future CDMA networks activated encryption schemes, however, SMS data would automatically be protected without any additional effort.

Sprint PCS and Nextel support one-way SMS only. In order to support two-way messaging, both networks provide WAP-enabled solutions. WAP provides security and authentication features similar to other Mobile Originated Messages, although the message is routed through the WAP server on the SS7 network, which adds an additional step in the message lifecycle.

Of all the technologies, GSM offers the most comprehensive protection for SMS data.

- Each GSM device contains a Subscriber Identity Module (SIM) card, a removable memory module with the subscriber's authentication key, personal identification number, and information for authentication and cipher key generation.
- Authentication is provided by a dedicated Authentication Center (AUC), which uses a challenge-response algorithm and shared secret data to authenticate the subscriber.
- The GSM A5 algorithm provides privacy by encrypting voice, data, and SMS traffic. Because A5 is a private key algorithm, the sender and receiver use the same key to encrypt and decrypt the data. To avoid transmitting these keys over the air, the MS and network use information from the authentication phase to generate them independently. The A5 encryption key is believed to be approximately 40 bits to 64 bits in length, meaning a machine testing one million keys per second would take between 13 days (40 bits) and

584,542 years (64 bits). A conservative estimate would be that a message is protected for several weeks.

- Finally, GSM provides anonymity by not transmitting a subscriber's telephone number on the air interface. This is implemented through the use of International Mobile Subscriber Identity (IMSI) and International Mobile Subscriber Identity (TMSI). A user's IMSI is transmitted when a handset is switched on; however, from that point forward, the network uses a series of ciphered, Temporary Mobile Station Identifiers (TMSIs) to identify each subscriber user. Even if an attacker decrypts the air interface and decodes some information, the identity of the cell phone user would still not be revealed.

TDMA employs an encryption scheme called the Cellular Message Encryption Algorithm (CMEA), a 64-bit block cipher encryption scheme. CMEA is considered less secure than GSM's A5 algorithm because a key vulnerability was discovered in 1997. Using this exploit, a well-equipped cryptographer could decode a message in less than a day. Although the CMEA algorithm is not appropriate for transmitting sensitive data, it does act as a strong deterrent to casual interception.

### **4.3 Writer-to-Reader Security**

It is important to understand that security features on the air interface and fixed network are independent. The network security features do not continue past the base station. This contrasts to the air interface which is a medium accessible by anyone. This has led to the different security models for each medium. Wireless networks were not designed to provide high security in a writer-to-reader sense and are not suitable for transporting sensitive or higher data.

The only alternative to protect data privacy in SMS messages involves encrypting the message body at the initial transmitter (writer) and then decrypting the message at the receiving device (reader). Such a method would protect the data for the entire duration of its transit through the network, although implementation of such a system is highly difficult. In order to be compliant with the all flavors of SMS, the encryption algorithm must possess three attributes; First, the algorithm cannot result in pure binary data. The encrypted message must be in the form of ciphertext in order to meet the SMS message body standards. Secondly, the encryption algorithm cannot drastically alter the size of the message, since that would cause initially large messages to exceed the maximum allowed size after encryption. Third, and most difficult, is that it cannot be a shared public key algorithm, as keys cannot easily be exchanged over the network. Private key algorithms are viable, as the cellular device can store a key, but this type of algorithm requires a difficult key management system that effectively prohibits its use.

Currently, however, the vast majority of cellular devices are phones without the ability to have new software installed on them. Individual programmers have managed to dismantle two models of Nokia GSM phones and rewrite the firmware as a proof of concept to allow for encrypted text messaging, but with no documentation on the phone's hardware or support from the manufacturers, this software is extremely unreliable and can never progress beyond what would be considered an alpha level. Furthermore, it suffers from the key management problem.

Finally, the very latest handset capabilities could greatly enhance SMS security. New devices such as the PocketPC Phone Edition and SmartPhone contain enough memory and computational power to run advanced cipher algorithms. Applications can be written on these devices to extract/capture short messages as they arrive at the handset. By combining these capabilities, it is possible to securely transmit sensitive information via SMS.



## 5. Conclusions

There is little doubt about the success of the Short Message Service across the globe among individual users. In Europe alone, messaging volumes are well over \$10 billion a month, with Asia adding huge volumes for a combined revenue realization of over \$25 billion worldwide at the end of 2002.

The reliability of wireless networks was examined. Generally speaking, wireless networks aspire to be as reliable as their wireline counterparts. Many mechanisms, such as fault tolerant and Network Equipment Building Standards (NEBS) compliant configurations, are in place to ensure high availability. The air interface represents a reliability challenge, but these problems have been studied and are well understood.

SMS security and vulnerabilities were also investigated. While there may be some potential for security-related abuses, no new and alarming issues were revealed, and the impression is that wireless operators are aware of the risks and rewards of offering these services. Some vulnerabilities exist on the air interface side; for example, a coordinated Radio Frequency (RF) jamming attack could render wireless networks inoperable in certain areas. However, this is not a new vulnerability, and most RF systems are susceptible to it.

Other services were briefly investigated to determine their relevance to this investigation:

- EMS and MMS were considered important because they are available to subscribers today and represent the evolution of SMS. However, the current use of these services is so low they are not considered to be an issue at this time. Of the two, MMS is considered most likely to be adopted widely. Its technical implementation is so different from SMS that a separate analysis is required to fully investigate the potential impact to the PSN.
- Cell Broadcast was considered because it represents another possible use of text messaging and can generate load on the network. However, cell broadcast services have not been widely deployed in North America. If cell broadcast were widely deployed a separate analysis is required to fully investigate the potential impact to the PSN. While cell broadcast is an interesting technology and may be suited for NS/EP applications, it does not appear likely to interfere with current NS/EP applications. The following points contributed to this conclusion:
  - CBS was designed to efficiently transfer messages to large numbers of users, and as such does not have the potential to consume network resources in the same way SMS does.
  - Furthermore, while cell broadcast centers do connect to the fixed part of the wireless network; they are not required to use SS7 system for message transfer.
  - Awareness and deployment of CBS is relatively low compared to SMS, meaning relatively little broadcast traffic is carried on wireless networks today.

Significant attention was given to network capacity. Networks are engineered for the loads they carry every day; they are not engineered for crisis situations. It is difficult to predict with certainty how a given network will behave when an extraordinary SMS load is applied to it.

Generally, it is within provider's network capability to continue operating and providing NS/EP services. However, most congestion algorithms are not addressed by specifications. The requirements each vendor uses when building congestion control algorithms are proprietary. Full certainty could only be obtained by joint investigation with each wireless operator and their equipment vendors.

The fixed network consists of SS7 message channels, short message centers, switching centers, and subscriber databases. Congestion may occur within any of these network elements. Each element can detect congestion, and in many cases relay information to connecting nodes for appropriate action. Multiple mechanisms handle congestion, with the general goal being to stem new traffic before it consumes network resources. These mechanisms can be effective on wireline networks, as past experience with the PSN has shown.



## 6. Recommendations

The investigation has concluded that under current deployments and the assumed usage patterns, SMS traffic will not impact the SS7 network and corresponding NS/EP applications. This is true if the anticipated load remains low and the analysis' assumptions on how the network handles congestion are correct. However, if these assumptions are not correct, high traffic load regions, such as Washington DC and New York, will experience significant traffic SMS congestion. The impact will be on the SMS users (denial of service) and not on the PSN or SS7. Any impact on NS/EP users depends on whether they are the ones trying to send/receive a message.

- Given the above, the NCS should undertake a more thorough analysis of potential traffic loads and congestion algorithms. Currently, coarse calculations point to problems when all subscribers in densely populated areas (Washington DC and New York City) attempt to use wireless networks. The accuracy of this assumption is uncertain. This analysis will require access to proprietary information from wireless equipment vendors and wireless network operators. Using this information, actual congestion situations could be identified, specific analyses could be performed for areas of concern to the NS/EP community, and activities begun to address assured usage for NS/EP users.
- The NCS should undertake a study to determine the feasibility of priority access for SMS messages. Similar to the WPS, the NCS should determine if the various SMS technologies and the features of SS7 can support the identification and granting of preferential treatment to NS/EP messages over non-NS/EP messages.
- The NCS should participate and provide contributions to the standardization groups that deal with future network architectures and features, such as 3GPP and 3GPP2. While the primary focus of these groups is the functionality provided to an average subscriber, the NCS should investigate network behavior to determine if standards could be specified to ensure the priority treatment outlined above can be implemented.



## Appendix A - SMS Operation

### A.1 Basic Principles

SMS allows users to send and receive short textual messages directly on their cellular or PCS cell phones. The screens below demonstrate how a subscriber sends, and receives, a short message. Figures A-1 and A-2 depict the generation of an SMS message on the two major wireless technology networks - Global System for Mobile Communications (GSM) network and Code Division Multiple Access (CDMA).



Figure A-1: The SMS Process on an Ericsson T68i in the USA T-Mobile Network



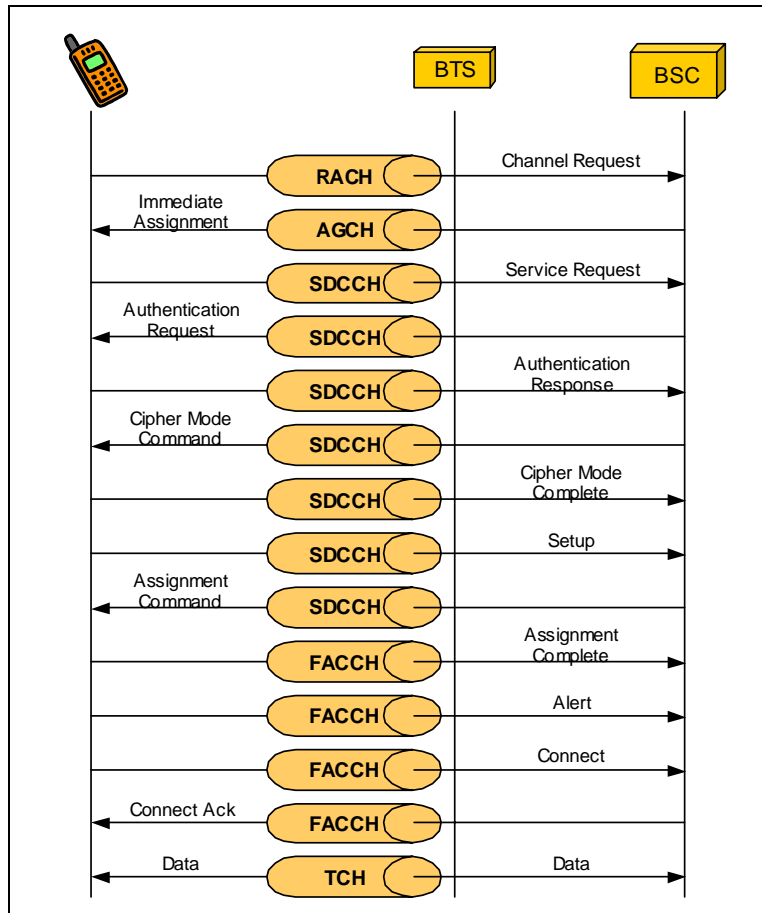
Figure A-2: The SMS Process on a Nokia 3285 in Verizon's Wireless Network

## A.2 Technical Discussion

### A.2.1 Mobile Originated Traffic

#### A.2.1.1 Mobile Originated Calls

Before a cell phone user can talk on their phone, a series of messages are sent between their phone and the network. Figure A-3 below shows the name of each message, its direction (phone to network, or network to phone), and GSM channel used.

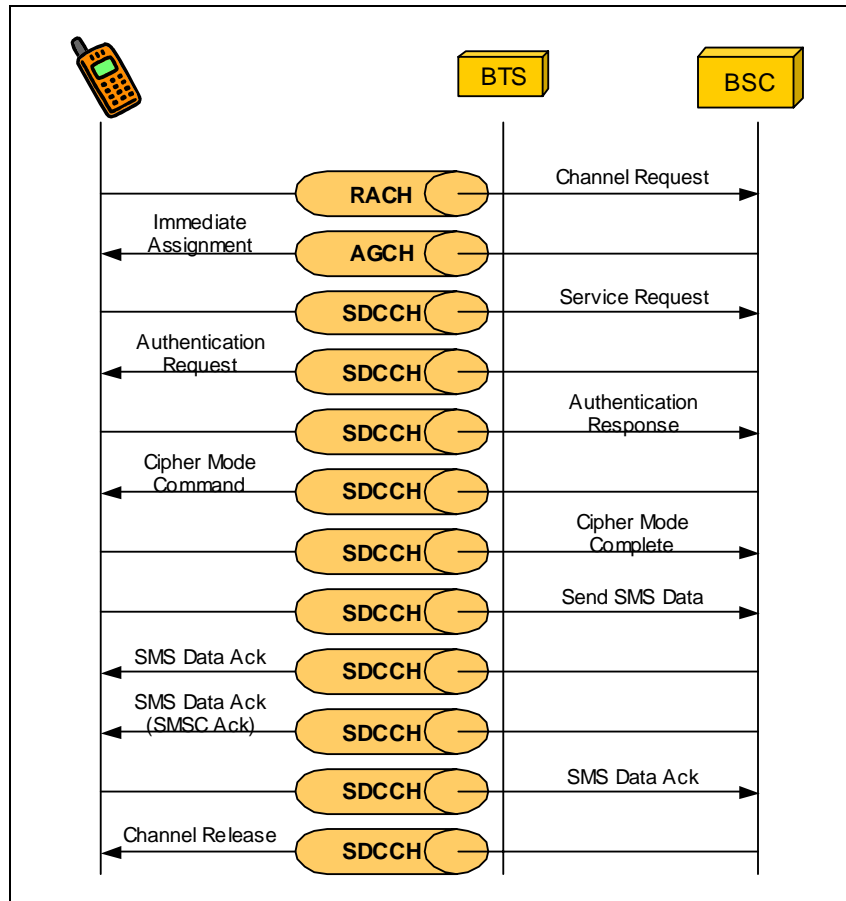


**Figure A-3: Air Interface Channels Used for Mobile Originated Voice Call**

The main channels used are the RACH, the SDCCH, and the TCH. The FACCH is also used, but is contained within a TCH, and for our purposes can be considered synonymous with the TCH.

#### A.2.1.2 Mobile Originated SMS

When a GSM cell phone sends an SMS, the process is similar to making a phone call. One major difference, however, is that an SMS does not require a traffic channel (TCH) to be allocated. Figure A-4 Depicts the Air Interface for SMS calls.



**Figure A-4: Air Interface Channels Used for Mobile Originated SMS**

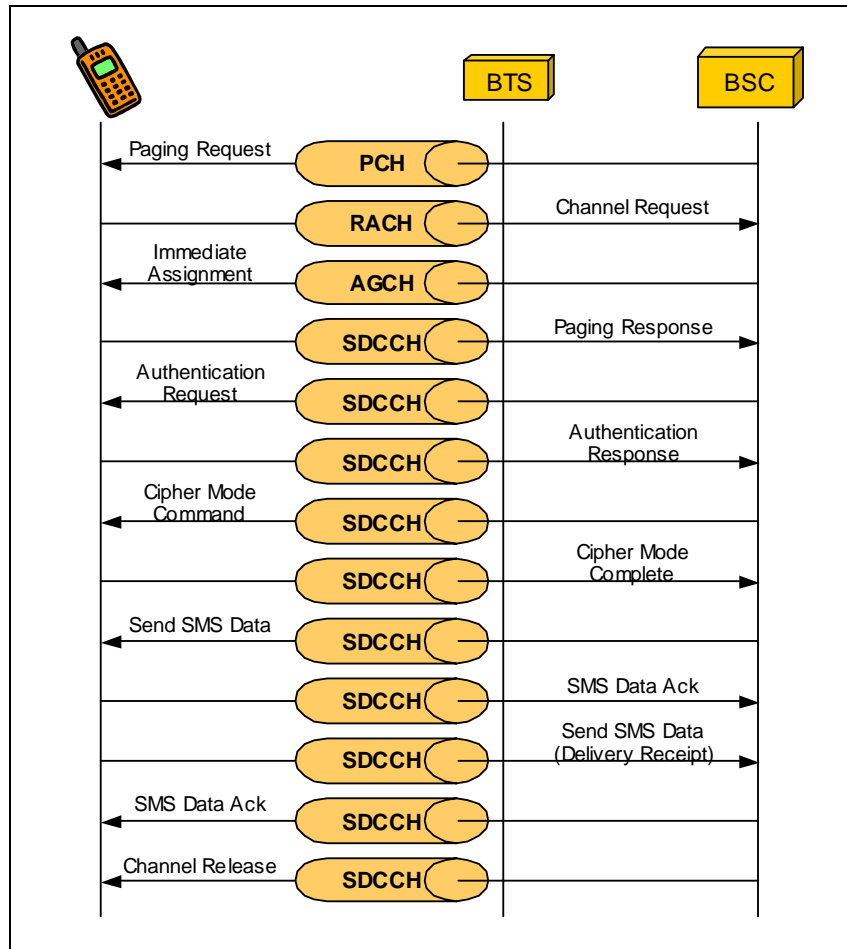
The main channels used are the RACH and the SDCCH.

## A.2.2 Mobile Terminated Traffic

### A.2.2.1 How a Cell Phone Receives an SMS

The main difference between cellular originated the traffic and cellular terminated traffic is the use of the Paging Channel (PCH). The paging channel notifies the cell phone of an incoming event, such as an arriving incoming voice call or an arriving SMS.

As shown in Figure A-5 when a subscriber receives a voice call or short message, the network sends the Paging Request on every PCH in the location area. A location area is a group of cell sites providing contiguous coverage across a geographic area. The network must track the user's Location Area at all times; when the cell phone moves from one location area to another, it informs the network of its new location. After receiving the page request message, the cell phone responds using the RACH of the closest cell site. Once the cell phone paging step is complete, call setup for cellular terminated traffic is similar to cellular originated traffic.

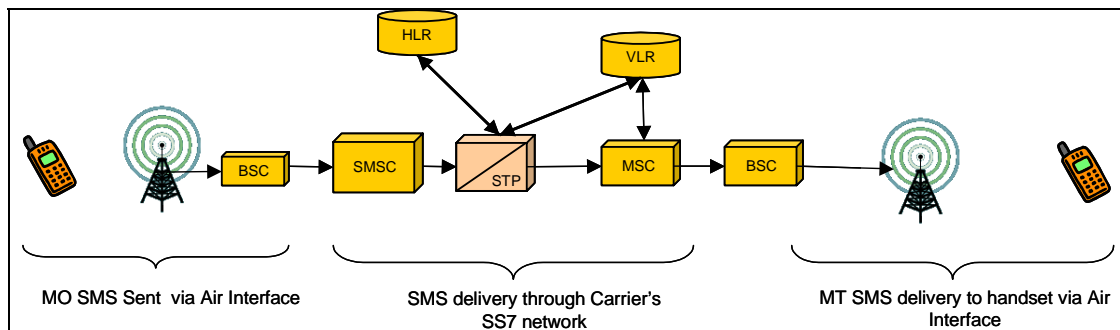


**Figure A-5: Air Interface Channels Used for Mobile Terminated SMS**

Short messages are sent several ways: from one cell phone to another, from a software program to a cell phone, or from a cell phone to a software program. How SMS messages traverse are discussed below in terms of the physical and functional pathways.

#### **A.2.2.2 Cell Phone-to-Cell Phone SMS on the Same Network**

Cell phone-to-cell phone messaging is the most popular form of messaging today. Subscribers can type in a message on their cellular SMS client and send it to another cell phone residing on the same system. Figure A-6 below shows how such a message traverses the network. The SMS message emanates from a users cell phone over the air interface. It is then sent through the Base Transceiver System (BTS), through the Base Station Controller (BSC – not used in CDMA), and then to the carrier’s SMS Center (SMSC). The SMSC queries the Home Location Register (HLR) via the Signaling System 7 (SS7) network, which consists of the Signaling Transfer Point (STP) and Visitor Location Register (VLR) residing in the Mobile Switching Center (MSC) where the target phone is active. The message is then sent through the MSC to a BSC, eventually arriving at the correct BTS and over the air interface to the target cell phone. A detailed description of each network component is provided in a later section.



**Figure A-6: Mobile-to-Mobile SMS**

In any SMS message transaction, the following are procedures that are invoked, along with the function names that are given for the procedures within both GSM MAP and IS-41 MAP protocols. [44]

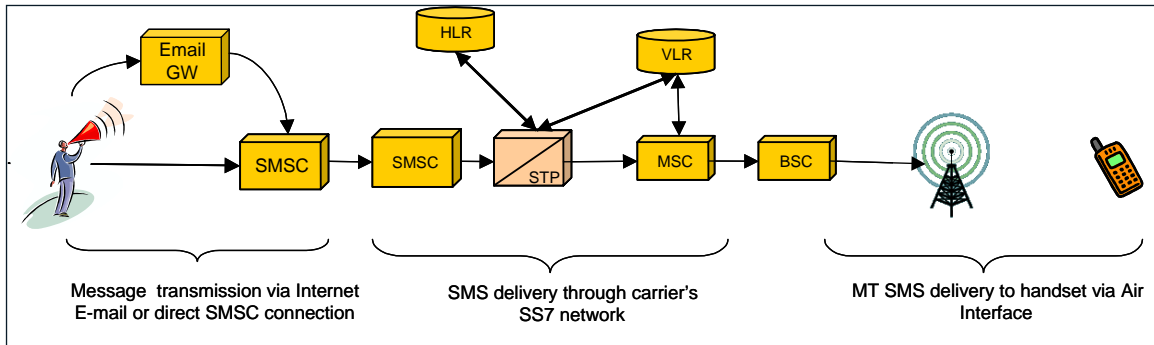
- Routing Information Request
  - IS-41: *SMSrequest*
  - GSM: *SendRoutingInfoForShortMsg*.
- Point-to-Point Short Message Delivery
  - IS-41: *short message delivery–point-to-point* (SMDPP)
  - GSM: *forwardShortMessage*
- Short Message Waiting Indication
  - IS-41: *SMS\_notification* indicator
  - GSM: *set\_message\_waiting\_data*
- Service Center Alert
  - IS-41: *SMS\_notification*
  - GSM: *alert\_service\_center*

The message processes shown in the figures below identify events that occur between network elements during SMS message transfer and are shown for both GSM and IS-41 network types. [45] While many delivery combinations exist, the focus will be on successful cellular originated and cellular terminated messages for GSM MAP and IS-41.

### A.2.2.3 Software Program to a Mobile Phone

Software programs can deliver messages to wireless networks in several ways. For example, networks allow messages to be sent as an email to a handset's phone number (e.g., (XXX)YYY-ZZZZ@messaging.sprintpcs.com). These emails arrive at a special Email Gateway, which converts them to SMS format and routes them to the network's SMSC. Alternatively, the software program can communicate directly with the SMSC using either the Internet or a direct

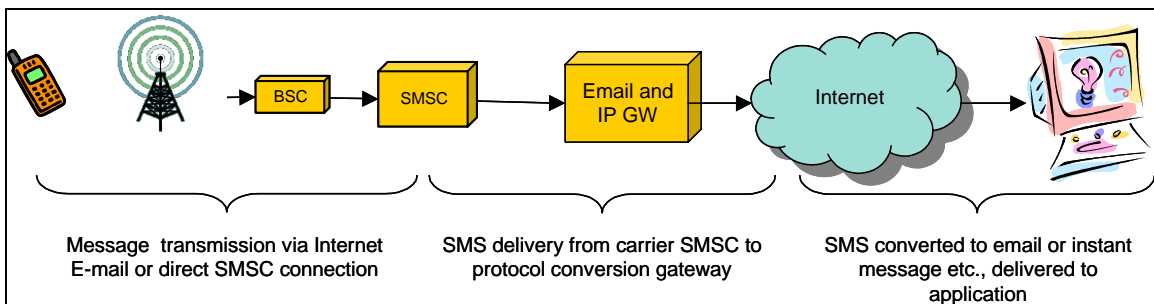
connection to send the message. The benefits and drawbacks of each approach will be considered in the Investigation section. Both paths are illustrated on the left side of Figure A-7 provided below.



**Figure A-7. Application-to-Mobile SMS**

#### A.2.2.4 Mobile Phone to a Software Program

For a cell phone to send SMS to an application, the application needs to have a gateway server sitting on the network with an address recognizable by the SMSC. The message is converted to an IP (Internet Protocol) based protocol such as email or instant messaging. Figure A.8 depicts the message flow.



**Figure A-8. Mobile-to-Application SMS**

### A.3 Special Case Handling

#### A.3.1 Phone Powered Off

The SMSC will continually monitor the network to identify when the phone is powered on. When the phone is powered on by a user, the SMSC registers that phone is active along with the MSC (location or city) where the phone is currently located. The registration notification that travels through the SS7 system to the wireless carriers HLR/VLR will also notify the SMSC that the device has been powered on, at which time the SMSC will take any stored messages and forward them to the mobile. If the phone is not powered on or registered within a certain time period, the SMSC will automatically delete messages from the buffer. The sender is notified on a case-by-case basis depending upon the carrier's operational setup.



### **A.3.2 Phone Out of SMSC Coverage Interoperability Area**

When the phone is out of the coverage area, it is treated as if it were powered off. If the phone is identified to be in a roaming partner's network that does not yet support SMS, the SMSC will retain the message for some time and wait for the phone to enter an SMS covered region. This still occurs when digital phones roam from their native digital systems onto an outdated, possibly rural, AMPS (Analog – Advanced Mobile Phone System) system.

### **A.3.3 Phone Registered in Multiple VLRs**

Multiple VLR registration occurs very infrequently with carriers. However, if a carrier has two MSC sharing the same boundary setup incorrectly in terms of the registration interval versus change of location area registration, the phone has the ability to toggle between the two systems, possibly showing up in the wrong VLR. Consequently the SMS would be delivered to the wrong MSC location and the user would never receive the message.

### **A.3.4 Phone Service Terminated by Customer**

When a recipient's service has been terminated, the message will not be sent and the user will not know if the intended subscriber has received it. Several operators have an auto-reply notifying the sender that the recipient has relinquished their service.

## **A.4 Other Message Combinations**

While the above diagrams indicate typical message interaction between elements on the same type of network, it should be noted that various combinations of SMS message origins and destinations can exist because of the different types of networks involved. Networks using IS-41 SS7 messaging may need to send messages to other networks using GSM SS7 messaging, and both of these types of networks may need to communicate with TCP/IP networks carrying email, instant messages, and web-originated messages. In order for such interprotocol message transactions to take place, specialized gateways that mate different protocols' function types with each other have been put into place. These gateways will continue to evolve as each type of network adds new features such as multimedia messaging, file and image transfer, etc.

Even if an SMS message is between two entities within a homogenous network, message transfer reliability issues do arise. The following are examples of issues that continue to affect SMS message delivery reliability [40]:

- Subscriber roaming into networks that do not support SMS
- Delivery during intersystem (intercarrier or interMSC) handoff
- Messages sent between systems with different maximum message lengths
- Delivery confirmation across gateways to other network types

Because of these types of issues as well as user-caused impediments to SMS delivery, an SMSC follows procedures to reattempt delivery when delivery has not been acknowledged by the

destination device. These procedures include setting a flag in the HLR to alert the SMSC when a subscriber's handset returns to an available state, at which point an SMSC will reattempt delivery. Furthermore, an SMSC may attempt delivery at regular intervals if no delivery confirmation has been received but the HLR shows the destination device to be available. Intervals between retries are normally set according to a wireless carrier's preference and may depend on local message volume and traffic handling capabilities.

Usage statistics on wireless networks show as many as 60% of subscribers may have their cell phones off at a given time. As a result, it is possible for an SMSC to build up a large queue of undelivered messages. In order to keep SMSC resources from having to grow uneconomically, a wireless carrier will normally set a message expiration time within the SMSC, typically on the order of several days or more. If a message cannot be delivered within the expiration timeframe, it is deleted from the queue, and its contents are no longer available.

## Appendix B - List of Acronyms

|         |  |
|---------|--|
| AMPS    | Advanced Mobile Phone System                           |
| AMR     | Adaptive Multi Rate                                    |
| ANSI    | American National Standards Institute                  |
| AT&T    | American Telephone and Telegraph                       |
| AUC     | Authentication Center                                  |
|         |  |
| BSC     | Base Station Controller                                |
| BSS     | Base Station Subsystem                                 |
| BSS-MSC | Base Station Subsystem-Mobile Switching Center         |
| BSSAP   | BSS Application Part                                   |
| BSSMAP  | Base Station Subsystem Management Application Sub-Part |
| BTS     | Base Transceiver System                                |
|         |  |
| CBC     | Cell Broadcast Centers                                 |
| CBS     | Cell Broadcast Service                                 |
| CDMA    | Code Division Multiple Access                          |
| CIMD    | Computer Interface to Message Distribution             |
| CMEA    | Cellular Message Encryption Algorithm                  |
| COTS    | Commercial Off the Shelf                               |
| CPU     | Central Processing Unit                                |
| CSC     | Common Short Codes                                     |
|         |  |
| DOS     | Denial of Service                                      |
| DTAP    | Direct Transfer Application Part                       |
|         |  |
| EMS     | Enhanced Messaging Service                             |
| ETS     | European Telecommunication Standards                   |
| ETSI    | European Telecommunication Standards Institute         |
|         |  |
| FACCH   | Fast Associated Control Channel                        |
| FEMA    | Federal Emergency Management Agency                    |
| FIPS    | Federal Information Processing Standard                |
|         |  |
| GIF     | Graphics Interchange Format                            |
| GPRS    | General Packet Radio Service                           |
| GSM     | Global System for Mobile communications                |
|         |  |
| HLR     | Home Location Register                                 |
|         |  |
| IEEE    | Institute of Electrical and Electronic Engineers       |
| IM      | Instant Messaging                                      |
| IMSI    | International Mobile Subscriber Identity               |
| IP      | Internet Protocol                                      |

|       |  |
|-------|--|
| IS-41 | Interim Standard 41                          |
| ISDN  | Integrated Services Digital Network          |
| JPEG  | Joint Photographic Experts Group             |
| KB    | Kilobyte                                     |
| L2ML  | Layer 2 Management Link                      |
| LAPD  | Link Access Procedure D                      |
| MAC   | Medium Access Control                        |
| MAP   | Mobile Application Part                      |
| MMS   | Multimedia Messaging Service                 |
| MMSC  | Multimedia Messaging Service Center          |
| MO    | Mobile Originated                            |
| MS    | Mobile Station                               |
| MSC   | Mobile Switching Center                      |
| MSN   | Microsoft Network                            |
| MT    | Mobile Terminal                              |
| MTP   | Message Transfer Part                        |
| MTP1  | Message Transfer Part 1                      |
| MTP2  | Message Transfer Part 2                      |
| MTP3  | Message Transfer Part 3                      |
| NCS   | National Communications System               |
| NEBS  | Network Equipment Building System            |
| NS/EP | National Security and Emergency Preparedness |
| OIS   | Open Interface Specification                 |
| SSL   | Secure Socket Layer                          |
| OML   | Operation and Maintenance Link               |
| OSI   | Open Systems Interconnection                 |
| P2P   | Peer to Peer                                 |
| PAD   | Packet Assembler Disassembler                |
| PC    | Personal Computer                            |
| PCH   | Paging Channel                               |
| PCIA  | Personal Communications Industry Association |
| PCN   | Personal Communications Network              |
| PCS   | Personal Communications System               |
| PDA   | Personal Digital Assistant                   |
| PDC   | Personal Digital Communications              |
| PSN   | Public Switched Network                      |
| PSTN  | Public Switched Telephone Network            |

|        |  |
|--------|--|
| RACH   | Random Access Channel                        |
| RF     | Radio Frequency                              |
| RSL    | Radio Signaling Link                         |
| SCCP   | Signaling Connection and Control Part        |
| SCP    | Service Control Points                       |
| SDCCH  | Standalone Dedicated Control Channel         |
| SHLR   | Standalone HLR                               |
| SIM    | Subscriber Identity Module                   |
| SMDPP  | Short Message Delivery Point to Point        |
| SMIL   | Synchronized Multimedia Integration Language |
| SMPP   | Short Message Peer to Peer                   |
| SMS    | Short Message Service                        |
| SMSC   | Short Message Service Center                 |
| SQL    | Structured Query Language                    |
| SS7    | Signaling System 7                           |
| SSH    | Secure Shell                                 |
| SSP    | Service Switching Points                     |
| STP    | Signaling Transfer Point                     |
| TAP    | Telocator Alphanumeric Protocol              |
| TCAP   | Transaction Capabilities Application Part    |
| TCH    | Traffic Channel                              |
| TCP    | Transport Control Protocol                   |
| TCP/IP | Transport Control Protocol/Internet Protocol |
| TDMA   | Time Division Multiple Access                |
| TIB    | Technical Information Bulletin               |
| TMSI   | Temporary Mobile Station Identifiers         |
| TRX    | Transceiver                                  |
| UCP    | Universal Computer Protocol                  |
| UDH    | User Data Header                             |
| UMTS   | Universal Mobile Telecommunications System   |
| U.S.   | United States                                |
| USA    | United States of America                     |
| VLR    | Visitor Location Register                    |
| W2F    | Wireless World Forum                         |
| WAP    | Wireless Application Protocol                |
| WPS    | Wireless Priority Service                    |
| WSP    | Wireless Session Protocol                    |



## Appendix C - List of References

- [1] Government of Canada Office of Critical Infrastructure Protection and Emergency Preparedness, *Incident Analysis: The September 11, 2001 Terrorist Attacks - Critical Infrastructure Protection Lessons Learned*, September 2002, [http://www.ocipep.gc.ca/opsprods/other/ia02-001\\_e.asp](http://www.ocipep.gc.ca/opsprods/other/ia02-001_e.asp)].
- [2] Alex LeVine and Sarah Roche, *Testimony for The Subcommittee on Science, Technology and Space*, December 5, 2001, <http://www.senate.gov/~commerce/hearings/120501Roche.pdf>].
- [3] [www.nokia.com](http://www.nokia.com)].
- [4] *Signaling System #7*, Travis Russel, McGraw-Hill, 2002.
- [5] [http://www.gcn.com/research\\_results/wireless-age4.html](http://www.gcn.com/research_results/wireless-age4.html)].
- [6] <http://www.pe.net/~rksnow/dc.htm>].
- [7] <http://www.wired.com/news/wireless/0,1382,55623,00.html>].
- [8] <http://www.demographia.com/db-nyc19102000.htm>].
- [9] <http://hypertextbook.com/facts/2002/JordanLevine1.shtml>].
- [10] Ertan Onur, Hakan Deliç, Cem Ersoy, and M. Ufuk Çağlayan *On the Retrial and Redial Phenomena in GSM Networks*, IEEE Wireless Communications and Networking Conference, September, 2000].
- [11] Prof. Raffaele Bolla,, [http://www.reti.dist.unige.it/telematica/lucidi/L5\\_Wireless\\_bw.pdf](http://www.reti.dist.unige.it/telematica/lucidi/L5_Wireless_bw.pdf), (11-21-2000)].
- [12] ETSI, *Digital cellular telecommunications system (Phase 2); Mobile radio interface layer 3 specification (GSM 04.08)*].
- [13] Barış Özgül, *Blind Collision Resolution for Wireless Multiple Access using Independent Component Analysis*, Master of Science in Electrical Engineering, Boğaziçi University, 2002].
- [14] Kyriazakos, Papaoulakis, Nikitopoulos, Gkroustiotis, Kechagias, Karambalis, and Karetos, *A Comprehensive Study and Performance Evaluation of Operational GSM and GPRS Systems under Varying Traffic Conditions*, Telecommunications Laboratory National Technical University of Athens, Greece.
- [15] ETS 300 595 *European digital cellular telecommunications system (phase 2); Base Station Controller - Base Transceiver Station (BSC - BTS) interface Layer 2 specification (GSM 08.56)*]

- [16] ETS 300 596 *European digital cellular telecommunications system (Phase 2); Base Station Controller - Base Transceiver Station (BSC - BTS) interface Layer 3 specification (GSM 08.58)*.
- [17] ETS 300 589 *European digital cellular telecommunications system (Phase 2); Signalling transport mechanism specification for the Base Station System - Mobile-services Switching Centre (BSS - MSC) interface (GSM 08.06)*].
- [18] ETSI *Digital cellular telecommunications system (Phase 2); Mobile radio interface layer 3 specification (GSM 04.08)*].
- [19] Giuseppe Bianchi - [http://www.tti.unipa.it/corsi/reti\\_radiomobili/2001-2002/slides/04-gsmair.pdf](http://www.tti.unipa.it/corsi/reti_radiomobili/2001-2002/slides/04-gsmair.pdf)].
- [20] EventHelix.com, *Network-wide Overload Control*:  
<http://www.eventhelix.com/RealtimeMantra/NetworkWideOverloadControl.htm>].
- [21] ETS 300 599 *Digital cellular telecommunications system (Phase 2); Mobile Application Part (MAP) specification (GSM 09.02 version 4.19.1)*.
- [22] Nagarajan, Ramesh, *Threshold-Based Congestion Control for the SS7 Signaling Network in the GSM Digital Cellular Network*, IEEE Transactions on Vehicular Technology, March 1999.
- [23] <http://www.mobileclb.com/services.asp>].
- [24] Gerrit Nieuwenhuis, *Citizen Alert via Cell Broadcast*, CGALIES 7th Plenary Meeting, July 2002, [http://cgalies.telefiles.de/Cell\\_Broadcast.zip](http://cgalies.telefiles.de/Cell_Broadcast.zip)}.
- [25] ETSI TS 101 368 V7.0.0 (1999-08), *Digital cellular telecommunications system (Phase 2+); Example protocol stacks for interconnecting Cell Broadcast Centre (CBC) and Base Station Controller (BSC) (GSM 03.49 version 7.0.0 Release 1998)*].
- [26] ETSI TS 123 041 V3.3.0 (2000-10) *Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); Technical realization of Cell Broadcast Service (CBS) (3GPP TS 23.041 version 3.3.0 Release 1999)*].
- [27] “Yes 2 EMS”, White Paper, Mobile Streams (September 2001)].
- [28] 3GPP 3<sup>rd</sup> Generation Partnership Project; Technical Specific Group Terminals; Technical realization of the Short Message Service (SMS) (Release 5.1.0) (2001-09)].
- [29] “What is EMS?”, magic4, (March 2003)].
- [30] GSM World, GSM Association (March 2003)].
- [31] Stephen Lawson, “SMS: Mobile Data’s Dark Horse Hits its Stride”, IDG News Service (March 2, 2001)].



- [32] SMS Forum, “Major Breakthrough for SMS Messaging,” Press Release (November 30, 2002)].
- [33] Wireless Internet Caucus, “Interoperable Text and Multimedia Messaging”, Cellular Telecommunications & Internet Association (2003)].
- [34] “USA – SMS Boom Set to Continue”, Mobile Youth (March 2003)].
- [35] W2F (Wireless World Forum), Mobile Youth, “USA – 450% Increase in SMS for Cingular Wireless” (January 31, 2002)].
- [36] Simon Buckingham, “Success 4 SMS” White Paper, Mobile Streams (February 2001)].
- [37] Cingular, “Business Benefits and Features”, Business Solutions (March 2003)].
- [38] Unimobile, “Enterprise Mobile Messaging Solutions” (March 2003)].
- [39] W2F (Wireless World Forum), Mobile Youth, “Government to Encourage Young with SMS Voting”, (February 7, 2002)].
- [40] *Mobile Telecommunications Networking with IS-41*, Michael D. Gallagher, Randall A. Snyder, McGraw-Hill, (1997).
- [41] Harry Newton, “Newton’s Telecom Dictionary, 17<sup>th</sup> edition”, CMP Books, (February 2001)
- [42] Telephony (Sept 17, 2001)