Abstract

This tutorial presents an overview of the Global System for Mobile Communications Short Message Service from the viewpoint of implementing new telematic services. SMS offers the users of GSM networks the ability to exchange alphanumeric messages up to the limit of 160 characters. The tutorial is motivated by an acute absence of research publications in this field. The information gathered in the tutorial was required considering the increasing potential SMS offers for integration with existing messaging services and its ability to offer a successful replacement for the Transmission Control and Internet Protocols as far as low-bandwidth-demanding applications are concerned. Initially, the tutorial gives a brief overview of the building blocks of GSM networks — the mobile station, base station, and network subsystem — and then emphasizes the SMS network and protocol architecture. The most widely used protocols for message submission are then introduced (text-based, SMS2000, ETSI 0705, TAP) and compared in terms of features provided and flexibility to handle extended alphabets or two-way messaging. Finally the tutorial outlines a summary of current and future issues for further development and research in the light of novel features for submission protocols and telematic services.

The Global System for Mobile Communications Short Message Service

GUILLAUME PEERSMAN AND SRBA CVETKOVIC,
THE UNIVERSITY OF SHEFFIELD
PAUL GRIFFITHS AND HUGH SPEAR, DIALOGUE COMMUNICATIONS LTD.

Since the first Global System for Mobile Communications (GSM) network started operation in 1991, more than 100 countries have adopted the standard. Over 20 million subscribers of GSM networks are now offered worldwide coverage, outstanding voice quality over a whole range of operating conditions, and a variety of value-added services. These services include voice mail, call handling facilities, call line identification, and Short Message Service (SMS).

With SMS, users are able to exchange alphanumeric messages (up to 160 characters) with other users of digital cellular networks, almost anywhere in the world, within seconds of submission. Even if the service was originally conceived as a paging mechanism for notifying the users of voicemail messages, SMS is now increasingly used as a messaging service. The messages are typically created on mobile phone keypads, which is somewhat awkward. Fortunately, there are other ways to access the message centers, as discussed in this article.

Numerous applications are already available and make short message reception and submission possible using a computer. Gateway architectures are also being widely implemented and connect company's e-mail or voicemail systems to the SMS.

The practical implementation of SMS and the different protocols for message submission are addressed in this article. The future of SMS and a brief review of the fields currently being studied will conclude this article.

The Short Message Service

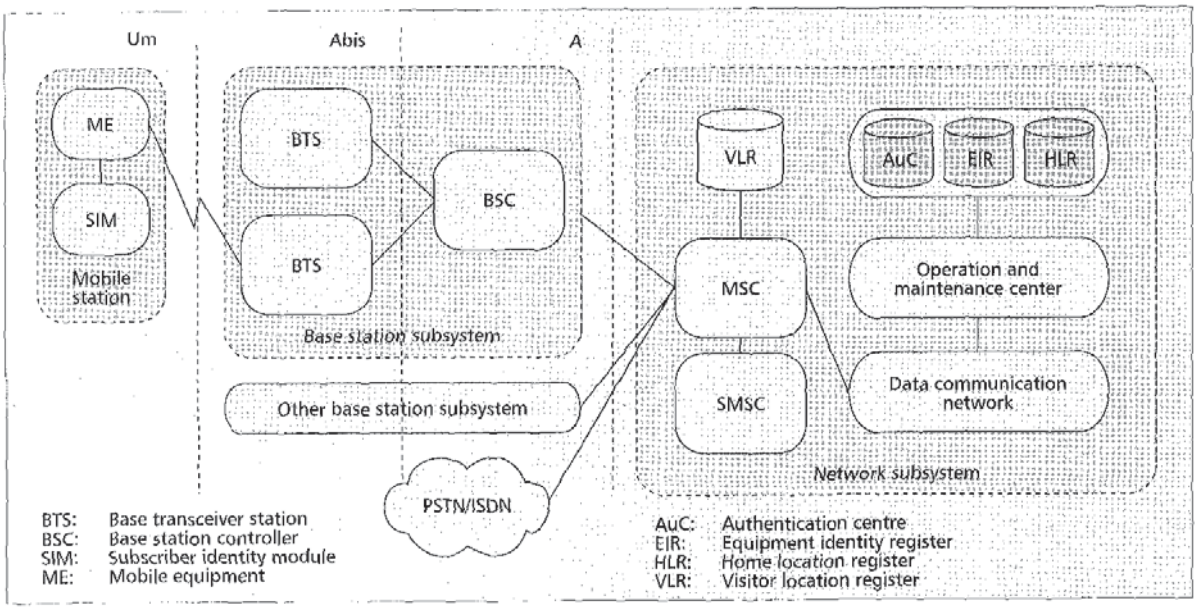
Developed as part of the GSM Phase 2 specification, the Short Message Service, or SMS as it is more commonly known, is based on the capability of a digital cellular terminal to send and/or receive alphanumeric messages. The short messages can be up to 140 bytes in length, and are delivered within a few seconds where GSM coverage is available. More than a common paging service, the delivery of the message is guaranteed even when the cellular terminal is unavailable (e.g., when it is switched off or outside the coverage area). The network will hold the message and deliver it shortly after the cellular terminal announces its presence on the network.

The fact that SMS (through GSM) supports international roaming with very low latency makes it particularly suitable for applications such as paging, e-mail, and voice mail notification, and messaging services for multiple users. However, the facilities offered to users and the charges for these facilities still mainly depend on the level of service provided by the network operator.

There are two types of SMS available: cell broadcast [1] and point-to-point [2]. In cell broadcast, a message is transmitted to all the active handsets or mobile stations (MSs) present in a cell that have the capability of receiving short messages and have subscribed to this particular information service. This service is only one-way, and no confirmation of receipt will be sent. It can send up to 93 7-bit character or 82 8-bit characters, typically used to transmit messages about traffic conditions, weather forecast, stock market, and so on.

In point-to-point service, messages can be sent from one mobile to another or from a PC to a mobile and vice versa. These messages are maintained and transmitted by an SMS Center (SMSC). The SMSC is an electronic form of ordinary mail postal service that stores and then forwards the messages when they can be delivered. Each GSM network must support one or more SMSCs to sort and route the messages. Each SMSC checks, organizes, and sends the message to the opera-

1070-9916/00/$10.00 © 2000 IEEE



■ **Figure 1.** *The basic GSM network architecture.*

tor. It also receives and passes on any confirmation messages to any GSM mobile on any network. However, in practice, there are no agreements to allow SMS to travel between networks.

There are several ways in which a short message can be submitted, depending on the interfaces supported by the GSM network SMSC. Users can call a central paging bureau (i.e., an operator), or directly create the message on the keypad of their handset. Typing the messages is made easier when using a personal digital assistant (PDA) or a laptop connected to the handset. A few SMSC equipment manufacturers and companies have also developed their own protocols for short message submission. Consequently, more and more GSM networks now offer access to their SMSC using these protocols over a variety of hardware interfaces: modem dialup, X25, and even the Internet.

GSM Network Architecture

The layout of a generic GSM network with its several functional entities is shown in Fig. 1 [3]. The architecture can be divided in three main components:

- The subscriber holds the MS, namely the GSM terminal
- The base station subsystem controls the radio link with the MS
- The network subsystem performs the switching of calls and other management tasks such as authentication.

The Mobile Station

The MS and base station subsystem communicate across the Um interface, also known as the air interface or radio link. The base station subsystem communicates with the network subsystem across the A interface. The MS consists of the physical terminal and contains the radio transceiver, the display and digital signal processors, and the Subscriber Identity Module (SIM). The SIM provides the user with the ability to access their subscribed services regardless of the location and the terminal used. The insertion of the SIM in any GSM cellular phone allows the user to access a network, make and receive phone calls, and use all the subscribed services.

The International Mobile Equipment Identity (IMEI) uniquely identifies the mobile terminal according to the International Mobile Subscriber Identity (IMSI) contained in the SIM. Because the IMEI and IMSI are independent, personal mobility is possible. The SIM can be protected against unauthorized use by a personal identity number (PIN).

The Base Station Subsystem

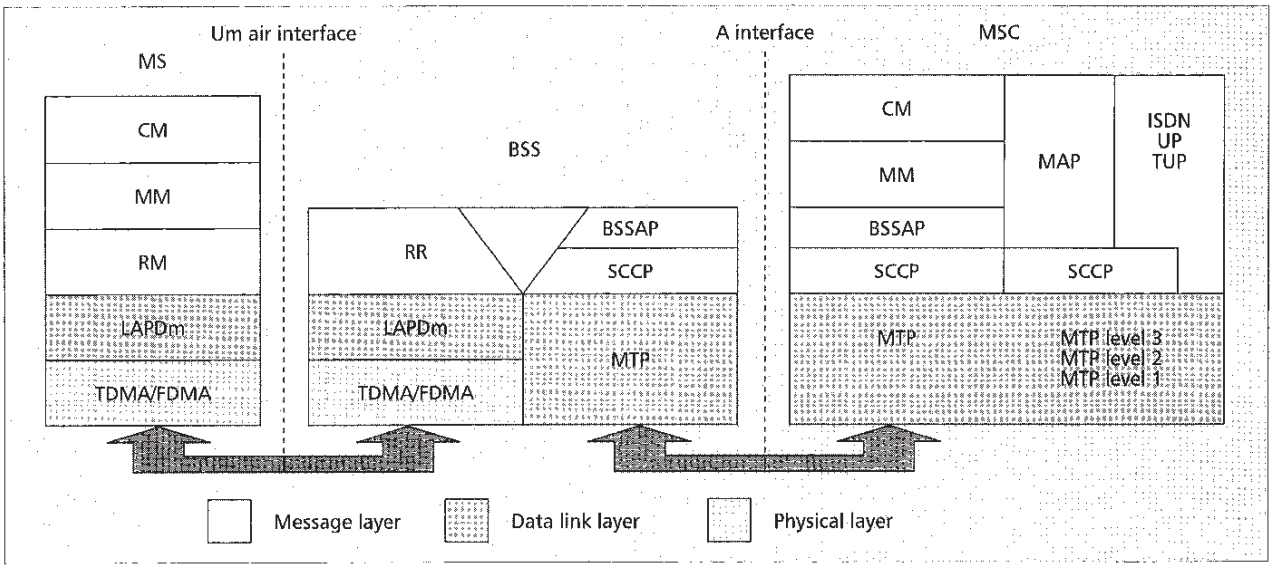
The base station subsystem is composed of two parts, the base transceiver station (BTS) and base station controller (BSC). They communicate across the specified Abis interface, thus allowing network operators to use components made by different suppliers. The BTS houses the radio transceivers that define a cell and handle the radio link protocols with the MS. Depending on the density of the area, more or fewer BTSs are needed to provide the appropriate capacity to the cell. Digital communications system (DCS) networks working at 1800 MHz need twice the number of BTSs to cover the same area as GSM networks, but provide twice the capacity.

The BSC manages the radio resources for one or more BTSs via the standardized Abis interface. It handles radio channel setup, frequency hopping, and handovers. The BSC is the connection between the MS and the mobile switching center (MSC). The BSC also takes care of converting the 13 kb/s voice channel used over the radio link (Um interface) to the standardized 64 kb/s channel used by the public switched telephone network (PSTN).

The Network Subsystem

The MSC is the main component of the network subsystem. Its provides the same functionality as a switching node in a PSTN or integrated services digital network (ISDN), but also takes care of all the functionality needed to handle a mobile subscriber such as registration, authentication, location updating, handovers, and routing to a roaming subscriber. The MSC also acts as a gateway to the PSTN or ISDN, and provides the interface to the SMSC.

The international roaming and call routing capabilities of GSM networks are provided by the home location register (HLR) and visitor location register (VLR) together with the MSC. The HLR database contains all the administrative information about each registered user of a GSM network along with the current location of the MS. The current location of an MS is in the form of a Mobile Station Roaming Number

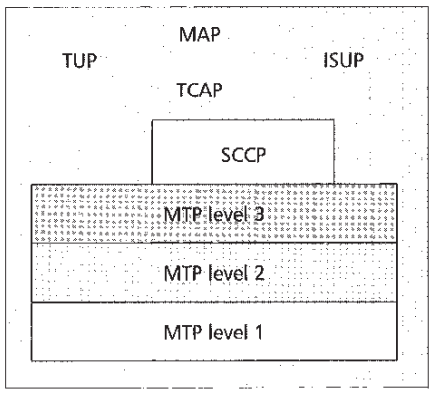


■ **Figure 2.** *The GSM protocol architecture.*

(MSRN), typically the SS7 number of the visited MSC, and used to route a call to the MSC where the mobile is actually located.

The VLR is usually located within the MSC to speed up access to the information required during a call and simplify the signaling. The content of the VLR is a selection of the information from the HLR, basically all necessary information for call control and provision of the subscribed services, for each single mobile currently located in the geographical area controlled by the VLR.

The network subsystem uses two other databases for authentication and security purposes. The Equipment Identity Register (EIR) contains a list of each MS IMEI allowed on the network. The authentication center (AuC) database contains each single PIN stored in the MS SIM.



■ **Figure 3.** *The SS7 protocol stack.*

The GSM Signaling Protocol

The exchange of signaling messages regarding mobility, radio resources, and connection management between the different entities of a GSM network is handled through the protocol architecture, as shown on Fig. 2.

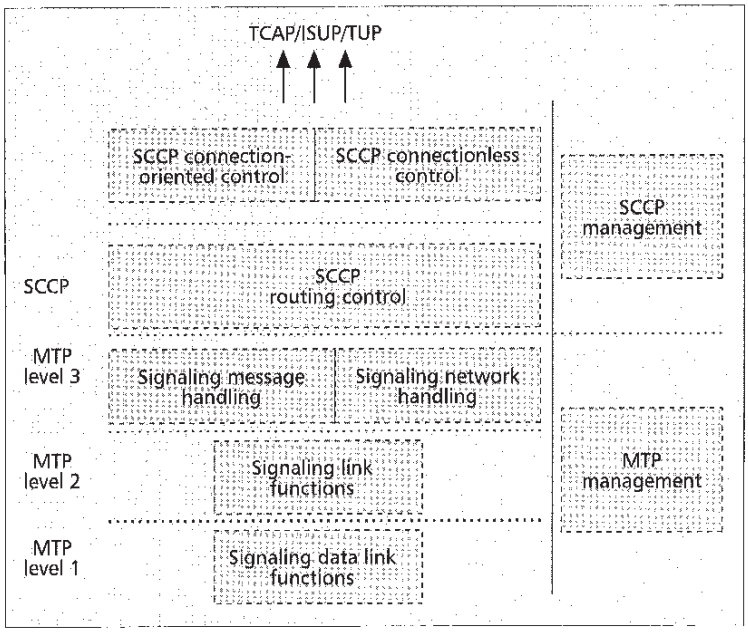
The architecture consists of three layers: physical, data link, and message. The physical layer and channel structure are described in detail by M. Mouly and M. Pautet [4]. Layer 2 implements the data link layer using a modified flavor of the Link Access Protocol (LAPD) to operate within the constraints set by the radio path. On the MS side, the message layer consists of three sublayers: connection management (CM), mobility management (MM), and resource management (RR). The CM sublayer manages call-related supplementary services, SMS, and call-independent supplementary services support. The MM sublayer provides functions to establish, maintain, and release a connection between the MS and the MSC, over which an instance of the CM sublayer can exchange information with its peer. It also performs location updating, IMSI management, and Temporary Mobile Subscriber Identity (TMSI) identification, authentication, and reallocation. The RR sublayer establishes the physical connection over the radio link to transmit call-related signaling information such as the establishment of the signaling and traffic channel between the MS and the BSS.

On the MSC side, the message layer is divided into four sublayers. The Base System Substation Application Part (BSSAP) of the MSC provides the channel switching functions, radio resources management, and internetworking functions. The Message Transfer Part (MTP) and Signaling Connection Control Part (SCCP) protocols are used to implement the data link layer and layer 3 transport functions for carrying the call control and mobility management signaling messages across the A interface. SCCP packets are also used to carry the messages for SMS.

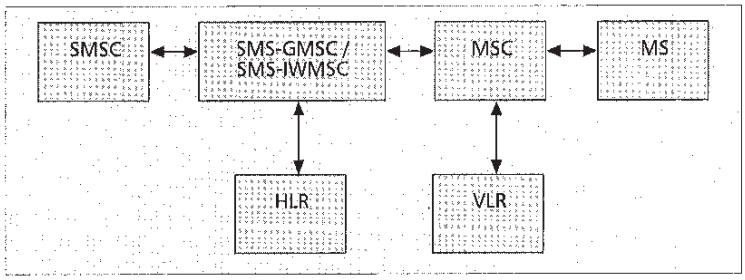
Signaling between the different entity uses the International Telecommunication Union (ITU) SS7, widely used in ISDN and current public networks. SS7 is currently the only element of the GSM infrastructure capable of packet switching as well as circuit switching. It is used to transport control signals and short message packets for SMS. The protocol consists of the Mobile Application Part (MAP), Transaction Capability Application Part (TCAP), SCCP, MTP, and ISDN-User Part (ISUP) or Telephone User Part (TUP). Figure 3 depicts the SS7 protocol stack.

The ISUP provides the signaling functions needed to support switched voice and data applications in the ISDN environment. The TUP provides the basic functionality for call control functions for ordinary national and international telephone calls. The TCAP is an application layer protocol. It allows an application at one node to invoke an execution of a procedure at another node and exchange the results of such invocation. It isolates the user application from the complexity of the transaction layer by automatically handling transaction and invocation state changes, and generating the abort or reject messages in full accordance with ITU and American National Standards Institute (ANSI) standards. The MAP uses the TCAP services to provide the signaling capabilities required to support the mobile capabilities.

The MTP and SCCP (Fig. 4) correspond to the lower three



■ **Figure 4.** *The SCCP and MTP sublayers.*



■ **Figure 5.** *The SMS network architecture.*

layers of the open system interconnection (OSI) model (Fig. 4). The SCCP sublayer supports connectionless and connection-oriented services to transfer data and Global Title Translation (GTT) above MTP level 3 for voice, data, ISDN, and GSM services. The data transfer is reliable, independent of the underlying hardware, and transparent to users. The protocol employs logical signaling connections within the SS7 network to ensure reliability and integrity of the ongoing data transfer. The MTP is divided into three levels:

- MTP level 1 defines the characteristics of the digital signaling link and is equivalent to the OSI physical layer.
- MTP level 2 is equivalent to the OSI data link layer and provides a reliable sequenced delivery of data packets across MTP level 1.
- MTP level 3 provides congestion control, signaling management, and message discrimination, distribution, and routing in a similar way as the OSI network layer.

Practical Implementation

SMS uses the SS7 signaling channel to transmit the data packet [5], thus allowing a text message to be received when the user is making a voice or data call. An active MS should be able to send and receive a short message Transport Protocol Data Unit (TPDU) at any time regardless of whether there is a speech or data call in progress. A confirmation will always be returned to the SMSC indicating whether the MS has received the short message or not. A confirmation will also be returned to the MS from an SMSC indicating whether the TPDU has been received successfully. The software within the MS must be able to decode and store the messages.

SMS Mobile Terminated (SMS-MT) is the ability to receive an SMS message from an SMSC and is more ubiquitous, while SMS Mobile Originated (SMS-MO) is the ability to send short messages to an SMSC. Messages can also be stored on the SIM, which can be retrieved at a later time. When the phone is not within coverage or the SIM is full, the SMSC will hold the message and deliver it shortly after the phone comes back into range or there is space in memory.

The SMS Basic Network Architecture

The main components of the SMS network architecture are shown in Fig. 5.

When routing a mobile originated short message, the SMSC forwards the short message to the SMS-GMSC. The SMS-GMSC interrogates the HLR for routing information and sends the short message to the appropriate MSC. The MSC delivers the short message to the MS. On the other hand, when routing a mobile terminated short message, the MS addresses the required SMSC according to its global title. If roaming abroad the visited public limited mobile network (PLMN) will route the short message to the appropriate SMS-IWMSC.

The SMSC identifies each short message uniquely by adding a time stamp in the SMS-DELIVER TP-SCTS field. The short message arrival at the SMSC is accurate to the second. It is the SMSC's responsibility to assure that if two or more short message arrive within the same second their time-stamps will be different.

The MS has to be able to receive/submit a short message TPDU, and then return a delivery report upon successful reception. It is also responsible for notifying the network when it has memory capacity available to receive one or more messages, if it had previously rejected a short message because its memory capacity was exceeded.

Protocol Architecture

The protocol layer for SMS is shown in Fig. 6. The short message transfer layer (SM-TL) services the short message application layer (SM-AL) and enables it to exchange short messages with a peer as well as receive confirmation of reception reports from earlier requests.

SMS-Deliver	Conveying a short message from the SMSC to the MS
SMS-Deliver-Report	Conveying a failure cause
SMS-Submit	Conveying a short message from the MS to the SMSC
SMS-Submit-Report	Conveying a failure cause
SMS-Status-Report	Conveying a status report from the SMSC to the MS
SMS-Command	Conveying a command from the MS to the SMSC

■ **Table 1.** *TPDU types.*

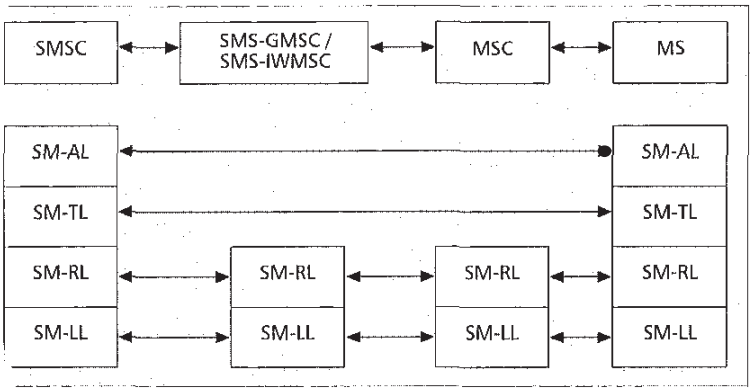


Figure 6. *The protocol layer for SMS point-to-point.*

The SM-TL exchanges PDUs with its peer entity. The short message relay layer (SM-RL) conveys the PDUs via the short message link layer (SM-LL). Refer to GSM 03.40 [2] for further details.

SMS Protocol Data Unit Types

There are six types of TPDU at the SM-TL, as listed in Table 1. The elements of the SMS-Deliver and SMS-Submit TPDU are shown in Fig. 7 [2]. The main fields of the TPDU are described in this document however for a complete description of the TPDU please refer to GSM 03.40 [2].

TP-Data-Coding-Scheme

The data coding scheme field (TP-DCS) is used to identify the coding scheme used by the user data, which can be 7- or 8-bit or even Unicode [6], as defined in GSM 03.38 [7].

TP-Validity-Period

The TP-VP field contains an information element enabling an MS to specify a validity period for the short message it is submitting. The value specifies how long an SMSC will guarantee the existence of a short message before delivery to the recipient has been carried out.

TP-More-Message-To-Send

The SMSC uses the TP-MMS field to inform the MS that one or more short messages are waiting to be delivered.

TP-User-Data-Header-Indicator

The 1-bit TP-UDHI field indicates whether the TP-UD includes an additional header as well as the short message.

TP-Protocol Identifier

The TP-PID is used by the MS or SMSC to identify the higher-layer protocol being used for internetworking with a certain type of telematic device (Telefax group 3 or 4, Ermes, etc.)

TP-User-Data (TP-UD)

The TP-UD field is used to carry the short message. It can store up to 140 octets of data for point-to-point SMS, together with a header depending on the setting of the TP-UDHI field. The amount of space taken by the header reduces the amount of data the PDU can carry. Figure 8 shows a representation of the layout of the TP-UD for 7- and 8-bit data schemes.

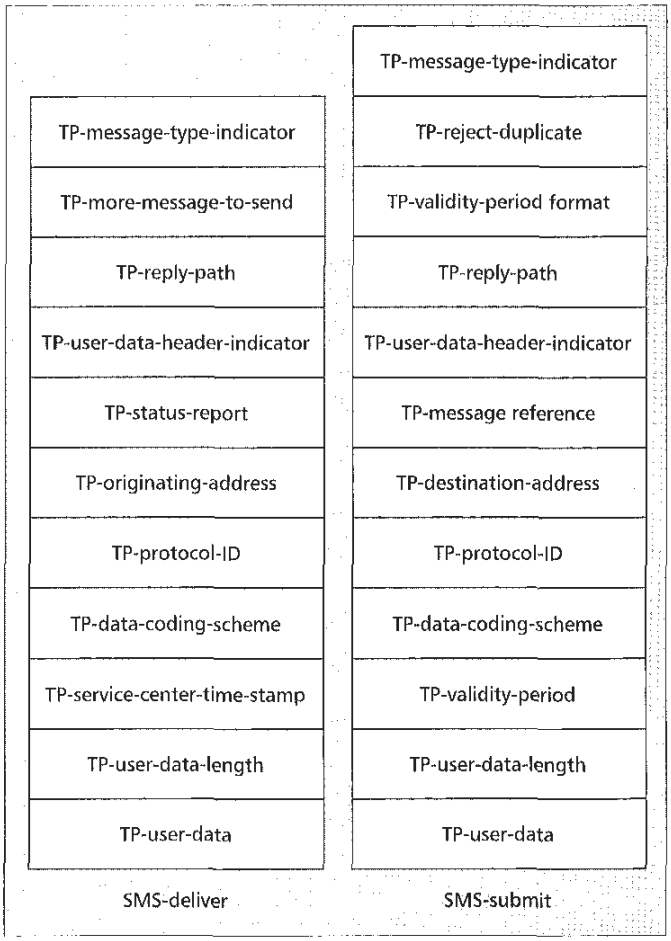
The header has at least three fields. The first field, the information element identifier, is used to identify concatenated short messages. Information data length (IDL) is used to indicate the length of the information element data (IED) that follows. Each of these fields is 1 octet long.

In the user data, the message can be 7 bits, 8 bits, or 16 bits. If 7-bit data is used and the header does not end on a 7-bit boundary, padding bits are used. This is to ensure that older mobiles which do not support the TP-UD header can still display the message properly.

Using the IEI allows sending and receiving of concatenated short messages. The IED field contains all the necessary information for the receiving entity to reassemble the messages in the correct order, and is coded as follows:

- First octet: short message reference number identifying the message within the same transaction
- Second octet: specifies the maximum number of short messages in the concatenated short message, which will not exceed 255
- Third octet: identifies the sequence number of the short message within the concatenated message

The minimum header length for concatenated message is 7 octets for 8-bit and 16-bit data and 8 for 7-bit data; leaving 133 (140 – 7), 152 (160 – 8), and 66 ((140 – 7)/2) characters for the short message. The maximum length of the message is then increased to 38,760 (255*152), 33,915 (255*133), or 16,830 (255*66) depending on the character coding scheme used.



■ **Figure 7.** *An SMS TL-PDU.*

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.