INNOVATIVE TECHNOLOGICA
METHODICAL RESEARCH JOURNAL
ISSN: 2776-0987 Volume 2, Issue 5, May, 2021

SEGMENT OPTIMIZATION OF SHORT MESSAGE SERVICE IN TELECOMMUNICATION

Rafidison Maminiaina Alphonse,
Telecommunication Automatic Signal Image Research Laboratory/
Doctoral School in Science and Technology of Engineering and Innovation/
University of Antananarivo, Madagascar

Ramafiarisona Hajasoa Malalatiana
Telecommunication Automatic Signal Image Research Laboratory/
Doctoral School in Science and Technology of Engineering and Innovation/
University of Antananarivo, Madagascar

Mr. Pravin R.Choube
Assistant Professor, Department of Electrical Engineering,
Dr. D.Y.Patil Institute of Technology, Pimpri, Pune, M.S., India

Abstract

Mobile Network Operator (MNO) subscribers are often surprised of Short Message Service (SMS) taxation because they ignore the number of part or segment of short message sent. The purpose of this research is to present a method to optimize the number of segments of SMS especially for multipart. A subscriber will be charged with minimum possible of segment. Commonly, GSM networks use as SMS encoding GSM-7 and UCS-2. The maximum character is very limited with UCS-2 encoding, so we will convert the encoding from UCS-2 to GSM-7 which supports more characters. It means that an application must be installed at end user Mobile Station (MS) which treats the original message before crossing GSM network. The called party has the same application to convert the receive message to the original one.

Keywords: Segment, SMS, Multipart, GSM-7, UCS-2.

Introduction

The Short Message Service (SMS) allows text messages to be sent and received to and from mobile telephones. The text can comprise words or numbers or an alphanumeric combination. SMS was created as part of the GSM Phase 1 standard. The first short

INNOVATIVE TECHNOLOGICA
METHODICAL RESEARCH JOURNAL
ISSN: 2776-0987 Volume 2, Issue 5, May, 2021

message is believed to have been sent in December 1992 from a PC to a mobile phone on the Vodafone GSM network in the UK. Each short message is up to 160 characters in length when Latin alphabets are used, and 70 characters in length when non-Latin alphabets such as Arabic and Chinese are used [1].

The maximum characters in one segment will be limited to 160 characters by ignoring the type of encoding used. To reach this goal, we will talk about SMS protocol stack, SMS encoding, the proposed algorithm followed by the conclusion.

1. SMS Protocol Stack

The protocol layers used in order to transmit an SMS can be seen in Fig. 1. The layers used during calls (Mobility Management MM, Radio Resource RR, Link Access Protocol on the Dm channel LAPDm, and Physical Layer) are also represented. The Connection Management CM layer includes a specific sublayer for the SMS.

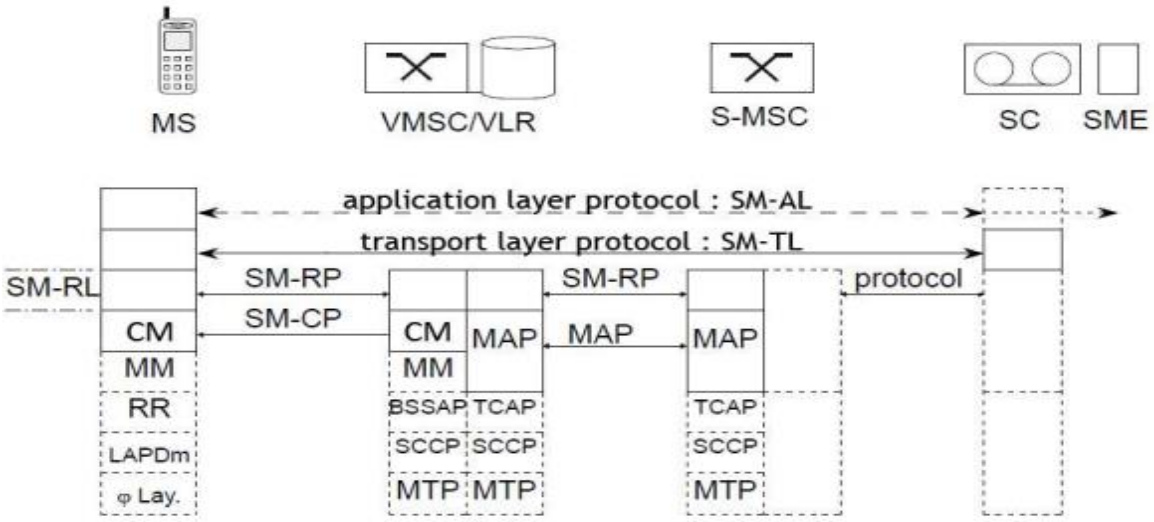

Fig. 1 Protocol implemented for SMS

There are 4 layers involved in the transmission of SMS:

- The Short Message Application Layer (SM-AL): is present in the mobile terminal and the Short Message Entity (SME). Its purpose is to generate and interpret messages.
- The Short Message Transfer Layer (SM-TL): offers the SM-AL a reliable service in order to transfer and receive short message between MS and Special Server (SC). It

INNOVATIVE TECHNOLOGICA
METHODICAL RESEARCH JOURNAL
ISSN: 2776-0987 Volume 2, Issue 5, May, 2021

performs the coding, and adds a timestamp of when the message was received by the server.

- Short Message Relay Layer (SM-RL): allows the transfer of messages through different equipment using Store and Forward. In GSM 2 protocols are specified for this layer:
  - Short Message Relay Protocol (SM-RP): it's used between the mobile terminal and the VMSC/VLR. It manages the references and addressing.
  - Mobile Application Protocol (MAP): it's used between the VMSC/VLR and the SMS-IWMSC.
  - The protocol between the SMS-GMSC/SMS-IWMSC and the SC does not concern GSM standards.
- Connection Management (CM): The Short Message Control Protocol (SM-CP) works between the MS and the VMSC/VLR. It transmits the SM, and protects against loss caused by changing the dedicated channel (this is needed because when changing the dedicated channel, the LAPDm demands a new one, and does not secure the transmission) [2].

2. SMS Encoding

Depending on which alphabet or character set you use, SMS messages typically contain a maximum of 160 7-bit characters or 70 2-byte characters. If the allowed character count is exceeded, the SMS is split into multiple messages and additional costs are assessed accordingly.

3.1 GSM-7

GSM-7 character set supports most, but not all, characters for languages that use the Latin-based alphabet, such as English, Spanish, French. The GSM character encoding uses 7 bits to represent each character similar to ASCII. One SMS message that uses GSM can contain a maximum of 160 characters.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z a b c d e f g h i j k l m n o p q r s t u v w
x y z 0 1 2 3 4 5 6 7 8 9 : ; < = > ? ¡ ¿ ! " # ¤ % & ( ) ' * + , - . / Ä Ö Ñ Ü § ä ö ñ ü à @ £ $ ¥ è é
ù ì ò Ç Ø ø Å å Δ _ Φ Γ Λ Ω Π Ψ Σ Θ Ξ Æ æ ß É

Fig. 2 GSM-7 characters

INNOVATIVE TECHNOLOGICA
METHODICAL RESEARCH JOURNAL
ISSN: 2776-0987 Volume 2, Issue 5, May, 2021

The following Characters are also supported by GSM 7 but these cost 2 GSM 7 Characters:

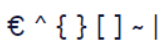

Fig. 3 Additional characters of GSM-7

GSM-7 messages break down as follows:

- Standard single SMS messages: For GSM phones with 7-bit character encoding, a standard SMS message can contain a maximum of 160 characters.

1120 bits / (7 bits/character) = 160 characters (01)

- GSM-7 multi-part or concatenated messages: When the message text is longer than 160 GSM characters, the message is concatenated and sent. When a message is concatenated, the user data header (UDH) consumes 6 bytes or 48 bits. This reduces the maximum number of characters in each message part:

1120 bits - 48 bits = 1072 bits (02)

1072 bits / (7 bits/character) = 153 characters per message part (03)

3.2 UCS-2

This character set is used for non-Latin based alphabet languages such as Arabic, Chinese, Cyrillic, and so on. As the characters for these languages are supported within Unicode, 16 bits per character is used instead of 7 bits per character [3].

The two types of UCS-2 messages are:

- Standard single SMS message (Unicode)- For Unicode phones with 16-bit character encoding, a standard SMS message can contain up to 70 characters. That is 1120 bits / (16 bits/character) = 70 characters.

1120 bits / (16 bits/character) = 70 characters (04)

INNOVATIVE TECHNOLOGICA
METHODICAL RESEARCH JOURNAL
ISSN: 2776-0987 Volume 2, Issue 5, May, 2021

- UCS-2 multi-part or concatenated messages- When the message text is longer than 70 UCS-2 characters, the message is concatenated and sent. When a message is concatenated, the user data header (UDH) consumes 6 bytes or 48 bits. This reduces the maximum number of characters in each message part:

1120 bits - 48 bits = 1072 bits (05)

1072 bits / (16 bits/character) = 67 characters per message part (06)

3. SMS Segment Optimization

Grace with two previous paragraphs, we are able to understand how a SMS transit in mobile network and know the structure data of SMS. The proposed method is shown in Fig.4. First of all, we extract the characteristic of original SMS such as encoding type and the length. If the type is GMS-7, we don't proceed to treat the SMS but in case it is UCS-2, we should consider the length. If the length is less than 70 characters, there is no processing same as the previous case because the SMS has only one segment, so no additional cost.

Once the SMS type is UCS-2 with characters more than 70, we swap some characters to GSM-7 without changing the signification of the short message. After this operation, if the type remains UCS-2, we convert the character to hexadecimal by adding "0x" as prefix. It will be required when we restore the original SMS.

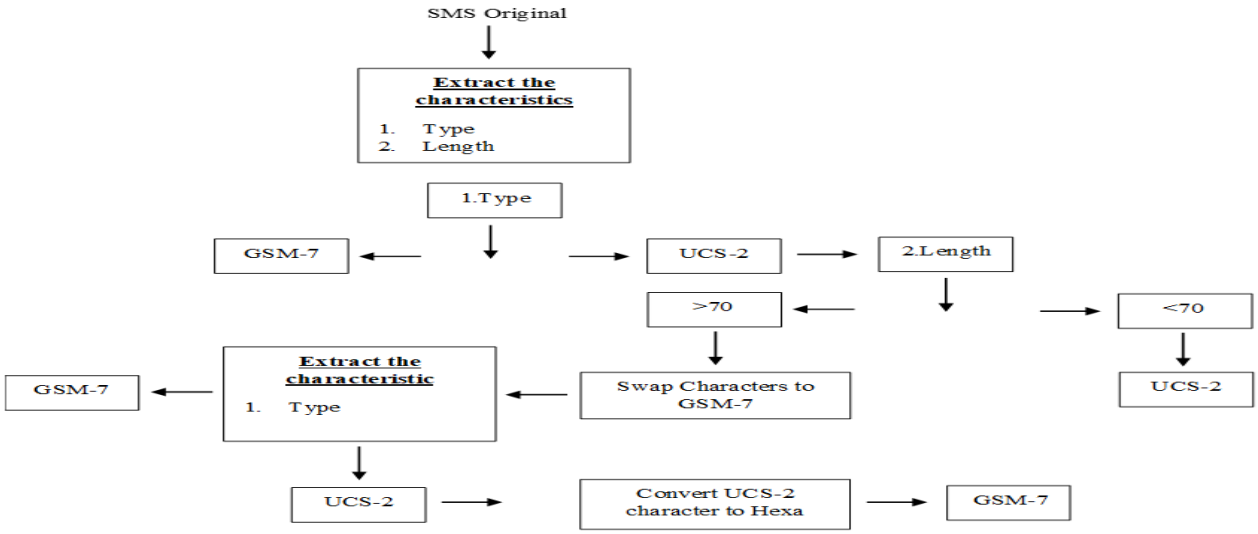

Fig. 4 SMS segment optimization algorithm

INNOVATIVE TECHNOLOGICA
METHODICAL RESEARCH JOURNAL
ISSN: 2776-0987 Volume 2, Issue 5, May, 2021

3.1Characteristic extraction

To extract the type and length, we set in table the list of GSM-7 characters (Fig. 2). We search each character of the original SMS in this table and calculate the total of number character found. If this number is less than the original SMS length, the SMS type is classified as UCS-2. Otherwise, it remains GSM-7.

According to Fig. 3, GSM-7 has an extra character coded with 2 bytes. We count these extra and add with original length to obtain the correct SMS length.

3.2Character swap

Character swap module is doing the mapping between UCS-2 and GSM-7 encoding as illustrated in Table 1. At the end of this step, if the UCS-2 character still exists, we convert them to hexadecimal.

Table 1: Mapping character

UCS-2	GSM-7	UCS-2	GSM-7	UCS-2	GSM-7
À	A	Ò	O	ç	c
Á	A	Ó	O	ê	e
Â	A	Ô	O	ë	e
Ã	A	Õ	O	í	i
Ä	A	Ö	O	î	i
Å	A	×	X	ï	i
È	E	Ø	0	ð	o
É	E	Ù	U	ó	o
Ê	E	Ú	U	ô	o
Ë	E	Û	U	ö	o
Ì	I	Ü	U	ú	u
Í	I	Ý	Y	û	u
Î	I	Þ	P	ý	y
Ï	I	á	a	þ	p
Đ	D	â	a	ÿ	y
Ñ	N	ã	a		

Experience

The original message has a GSM-7, UCS-2 and GSM-7 characters. The initial segment was 3 and reduced to 2 after conversion with GSM-7 encoding type.

3.3Original SMS

©[#Voy al mercado con mamá | #Je vais au marché avec maman | #Andeha ho any an-tsena hiaraka amin'i mama aho | #I go to the market with my mom | #Vado al mercato con mamma]'

INNOVATIVE TECHNOLOGICA
METHODICAL RESEARCH JOURNAL
ISSN: 2776-0987 Volume 2, Issue 5, May, 2021

Length: 180
Segment: 3
Encoding Type: UCS-2

3.4 SMS after swap

' ©[#Voy al mercado con mama | #Je vais au marché avec maman | #Andeha ho any an-tsena hiaraka amin'i mama aho | #I go to the market with my mom | #Vado al mercato con mamma]'

Length: 180
Segment: 3
Encoding Type: UCS-2

3.5SMS after Hexadecimal conversion

' 0xA9[#Voy al mercado con mama | #Je vais au marché avec maman | #Andeha ho any an-tsena hiaraka amin'i mama aho | #I go to the market with my mom | #Vado al mercato con mamma]'

Length: 183
Segment: 2
Encoding Type: GSM-7

3.6SMS received after decoding

' ©[#Voy al mercado con mamá | #Je vais au marché avec maman | #Andeha ho any an-tsena hiaraka amin'i mama aho | #I go to the market with my mom | #Vado al mercato con mamma]'

Length: 180
Segment: 3
Encoding Type: UCS-2

Conclusion

In this paper, we proposed a method to avoid an additional cost of SMS by finding a way to minimize the number of SMS segment. UCS-2 has limited characters due of 2

ISSN: 2776-0987 Volume 2, Issue 5, May, 2021

bytes encoding, so the idea was to translate UCS-2 characters to GSM-7 to have more characters in one SMS segment. A research was published in "International Journal of Trend in Scientific Research and Development (IJTSRD)" entitled "Image Encoding to Short/Text Message" to convert an image to short message and the initiative is from this approach. Image to text conversion may be has a long length and our method can be used to reduce the cost and the segment when sending it through GSM mobile network.

References

1. https://www.smssolutions.net/tutorials/misc/intro/
2. D. Ciuraru, L. Hilohi, A. Mercier, X. Lagrange, "SMS over LTE: services, architecture and protocols", 2013, pp.62. ffhal-00814264ff
3. R. Moore, "SMS message length and character encoding", https://help.goacoustic.com