

*Орешек знаний тверд, но...
Мы не привыкли отступать!
Нам расколоть его поможет
Киножурнал «Хочу все знать!»*

Алексей Печерский,
SIA «AP System»,

Рига, февраль 2008

Формат PDU или как сформировать и распаковать SMS

Предисловие

Предлагаемый вашему вниманию документ представляет собой попытку упорядочить информацию о формате PDU, принятом для передачи коротких сообщений в мобильных сетях. Большая часть материала проверена с помощью мобильного телефона SonyEricsson K750i и программы на Delphi 6.

Второй документ «Как послать и получить SMS или система GSM AT-команд» посвящен AT-командам.

Все допущенные в изложении ошибки лежат на моей совести.
Все конструктивные замечания и исправления приветствуются.

Оглавление :

1.	Общие положения	3
2.	Словарь.....	3
3.	Представление данных.....	4
4.	Взаимодействие компьютера и MS	5
5.	Типы PDU:	5
5.1.	Состав PDU	5
5.2.	Формат передаваемого SMS	6
5.3.	Формат принятого SMS	7
6.	SCA - Service Center Address - адрес сервисного центра рассылки коротких сообщений	8
7.	Protocol Data Unit Type (PDU type) - тип пакета данных протокола	10
7.1.	PDU type для SMS-Submit (Mobile Originated) - тип PDU для передаваемого SMS	10
7.2.	PDU type SMS-Deliver (Mobile Terminated) - тип PDU для принятого SMS	12
7.3.	Message Reference	13
7.4.	Originator Address OA и Destination Address DA - адрес источника и приемника	13
7.5.	Protocol Identifier PID - идентификатор протокола	13
7.6.	Data Coding Scheme DCS - схема кодирования данных	14
7.7.	Message Class - Классы сообщений	16
7.8.	Service Center Time Stamp SCTS - метка времени Short Message Service Center (центра обслуживания коротких сообщений)	17
7.9.	Validity Period VP - Время жизни сообщения	17
7.10.	User Data Length UDL - длина данных пользователя	18
7.11.	User Data UD - данные пользователя	18
7.12.	Метод упаковки данных пользователя	18
8.	Заключение	20
Приложение 1.		
	Таблица символов алфавита GSM по умолчанию (GSM default alphabet):	21
Приложение 2.		
	Таблица символов UCS2 - C0 controls and Basic Latin Range 0000-007F	22
Приложение 3.		
	Таблица символов UCS2 Cyrillic Range 0400-04FF	23
	Список литературы:	24

1. Общие положения

Согласно GSM 03.40 version 6.0.0 сервис коротких сообщений (Point-to-Point Short Message Service или, сокращенно, SMS) предоставляет возможности отправки и приема сообщений ограниченного размера мобильными телефонами. Эта возможность обеспечивается при помощи SMSC - Short Message Service Center (сервисный центр коротких сообщений, центр обслуживания коротких сообщений), который обеспечивает хранение принятых сообщений и дальнейшую их передачу адресатам. В свою очередь, в сервисе SMS определены два сервиса типа точка-точка: сервис по доставке сообщения от мобильного телефона до Сервисного Центра Коротких сообщений (Mobile originated) и сервис по доставке сообщения от Сервисного Центра Коротких сообщений до мобильного телефона (Mobile terminated).

Таким образом, отправленное сообщение сперва принимается оборудованием Сервисного Центра, который затем пытается передать сообщение адресату. Если доставка невозможна, например, телефон адресата вне зоны действия сети, Сервисный Центр сохраняет сообщение и продолжает попытки доставить сообщение адресату. Если сообщение доставлено или его время жизни истекло, оно удаляется.

Сервисы базируются на обмене PDU (Protocol Data Units или Packet Data Unit, пакетов данных протокола). PDU имеют бинарное представление для повышения эффективности передачи. Обмен пакетами может происходить как синхронно (после отправки запроса дальнейший обмен пакетами приостанавливается до получения ответа), так и асинхронно (запросы отправляются без задержек, обработка ответов происходит по мере их поступления).

Процесс передачи PDU скрыт от пользователя мобильной сети и для отправки сообщения достаточно правильно сформировать PDU (пакет данных протокола) и при помощи AT команд отправить их мобильному оборудованию (ME - телефону или модему), все остальное произойдет без участия пользователя. Аналогично происходит и прием сообщений - необходимо при помощи AT команд прочитать PDU и распаковать его.

Передачу SMS обеспечивает PDU типа **SMS-Deliver** (Mobile Terminated), прием SMS обеспечивает PDU типа **SMS-Submit** (Mobile Originated). На этих двух типах PDU мы и остановимся.

2. Словарь

MS	Mobile Station	мобильная станция - телефон или модем
ME	Mobile Equipment	мобильное оборудование
TE	Terminal Equipment	мобильный терминал
SMSC	Short Message Service Center	сервисный центр коротких сообщений, центр обслуживания коротких сообщений
MMI	Man Machine Interface	интерфейс «человек-машина»
PDU	Protocol Data Unit	пакет данных протокола
SM-AL	Short Message Application Layer	протокол приложений для коротких сообщений
SM-TL	Short Message Transport Layer	транспортный протокол для коротких сообщений
PDU Type	Protocol Data Unit Type	тип пакет данных протокола
TPDU	Transport Protocol Data Unit	пакет данных транспортного протокола
SCA	Service Centre Address	адрес сервисного центра - центра обслуживания

		коротких сообщений
OA	Originator Address	адрес источника (создателя)
DA	Destination Address	адрес приемника (получателя)
PID	Protocol Identifier	идентификатор протокола
DCS	Data Coding Scheme	схема кодирования данных
SCTS	Service Center Time Stamp	метка времени SMSC (центра обслуживания коротких сообщений)
VP	Validity Period	время жизни сообщения
UDL	User Data Length	длина данных пользователя
UD	User Data	данные пользователя
RP	Reply Path	запрос ответа от стороны, принимающей сообщение.
UDHI	User Data Header Indicator	заголовок данных пользователя
SRI	Status Report Indication	индикация запроса на получение отчета
SRR	Status Report Request	запрос на получение отчета
VPF	Validity Period Format	формат представления параметра «время жизни сообщения»
MMS	More Messages to Send	ожидание продолжения сообщения - «еще сообщения для отправления»
RD	Reject Duplicate	запрос на удаление одинаковых сообщений
MTI	Message Type Indicator	индикатор типа сообщения
SIM	Subscriber Identity Modul	SIM - модуль идентификации пользователя (подписчика)
MO PDU	Mobile Originated PDU	PDU для SMS, которое будет отправлено от MS к SMSC - передаваемое SMS
MT PDU	Mobile Terminated PDU	PDU для SMS, которое будет отправлено от SMSC к MS - полученное SMS

3. Представление данных

Представление данных в PDU происходит от нескольких методов шестнадцатеричного представления данных - «чистых» шестнадцатеричных чисел, десятичных чисел и символьного представления чисел. Так, к примеру, некоторые поля PDU сжаты в байты. Один байт в этом случае - это 2 шестнадцатеричные цифры.

Во всех документах, связанных с передачей данных, применяются термины октет и полуоктет для 8 и 4 бит соответственно. Это объясняется тем, что в теории передачи данных принято данные передавать не байтами, а октетами. Мы же будем применять «байт», имея в виду, что «байт» это октет бит, проще говоря, у нас байт - это 8 бит. В документе (2) различие между байтом и октетом, несмотря на то, что это и в том и другом случае 8 бит, объясняется следующим образом: «байты сохраняются в памяти, а октеты - передаются», то есть то, что в памяти - это байт, а то, что в эфире - это октет. Но нас интересует память, так что термин «байт» вполне приемлем.

Иногда данные представляются как десятичные цифры, тогда для представления одной десятичной цифры достаточно 4 бит (полубайт применяется везде в этих

документах, мы будем применять термин «тетрада», имея ввиду, что тетрада это 4 бита или половина байта).

Символьное представление базируется на документе GSM 03.38. Это представление применяется главным образом в User Data (UD) – Данных пользователя.

В документе приняты следующие обозначения для представления чисел:

- b – двоичное представление;
- d – десятичное представление;
- h – шестнадцатеричное представление.

То есть 00001010b = 10d = 0Ah.

4. Взаимодействие компьютера и MS

При взаимодействии компьютера и MS (мобильного телефона или модема), соединенных по RS-232, управление MS базируется на наборе AT-команд MS.

Все действия по передаче/приему SMS проводятся в соответствии с транспортным протоколом (SM-TL) и обеспечивают для протокола приложений операции по приему, передаче SMS и получению отчетов о ранее переданных SMS. Транспортный протокол (SM-TL) выполняет операции обмена информацией при помощи шести разных типов PDU, которые перечислены ниже. Мы рассмотрим далее только 2 из них – **SMS-Deliver** и **SMS-Submit**.

5. Типы PDU:

1. **SMS-Deliver** (Mobile Terminated) – тип PDU, обеспечивающий передачу сообщения от SMSC до MS (мобильного телефона получателя). По-человечески – протокол, обеспечивающий получение сообщений;
2. **SMS-Deliver-Report** – тип PDU, обеспечивающий передачу сообщения о причине отказа передачи сообщения от SMSC до MS (если необходимо);
3. **SMS-Submit** (Mobile Originated) – тип PDU, обеспечивающий передачу сообщения от MS (мобильного телефона отправителя) до SMSC. По-человечески – протокол, обеспечивающий передачу сообщений;
4. **SMS_Submit-Report** – тип PDU, обеспечивающий передачу сообщения о причине отказа передачи сообщения от MS до SMSC (если необходимо);
5. **SMS-Status-Report** – тип PDU, обеспечивающий передачу состояния от SMSC до MS (мобильного телефона получателя);
6. **SMS-Command** – тип PDU, обеспечивающий передачу команд от MS до SMSC;

5.1. Состав PDU

В общем виде PDU состоит из **SCA** и **TPDU**

PDU = SCA + TPDU, где

SCA – Service Centre Address – адрес сервисного центра рассылки коротких сообщений;

TPDU – Transport Protocol Data Unit – пакет данных транспортного протокола.

Формат поля SCA будет рассмотрен ниже.

Поле TPDU состоит из заголовка, содержащего информацию о содержании PDU и пакет данных.

Некоторые модели мобильных телефонов и GSM модемов не поддерживают полный формат PDU и могут работать только с форматом TPDU. В этом случае SCA берется из памяти SIM-карты, а поле SCA заменяется на 00h. Далее мы будем рассматривать случай полного PDU, состоящего из SCA и TPDU.

5.2. Формат передаваемого SMS SMS-Submit (Mobile Originated)

Обозначение	SCA	PDU type	MR	DA	PID	DCS	VP	UDL	UD
Размер, байты	1-12	1	1	2-12	1	1	0,1 или 7	1	0-140

Длина в байтах	Пояснение	Поле	Обозначение
1..12	12 байт максимум, может быть меньше. Это зависит от длины номера SMSC или вообще может быть 00h – тогда номер SMSC берется из настроек телефона (Sim-карты)	SCA	Service Center Address, см.1.2
1	один байт, в котором задается тип PDU – состоит в свою очередь из одно и двухбитовых полей. Именно битами в PDU Type различаются форматы передаваемых и принимаемых сообщений. В принимаемых сообщениях последние два бита этого байта равны 00b, а в передаваемых равны 01b – в нашем случае 01b	PDU Type	Protocol Data Unit Type
1	количество успешно переданных – от 00h до ffh – устанавливается самим телефоном – поэтому его надо установить при передаче в 00h	MR	Message Reference
2..12	Адрес (номер телефона) приемника сообщения или, попросту, кому посылаем SMS, телефон получателя	DA	Destination Address
1	Идентификатор протокола – указывает SMSC, как обрабатывать передаваемое сообщение (как факс, голосовое сообщение и т.д.)	PID	Protocol Identifier
1	Схема кодирования данных в поле UD	DCS	Data Coding Scheme
0,1 или 7	Время жизни сообщения –	VP	Validity Period

	указывает SMSC время, по истечении которого сообщение недействительно, попросте – сообщение уничтожается, если по истечении этого времени оно не доставлено (то есть не получено получателем – SMSC пыталась его передать, но получатель его не получил).		
0..140	140 байт данных пользователя	UD	User Data

5.3. Формат принятого SMS SMS-Deliver (Mobile Terminated)

Обозначение	SCA	PDU type	OA	PID	DCS	SCTS	UDL	UD
Размер, байты	1-12	1	2-12	1	1	7	1	0-140

Длина в байтах	Пояснение	Поле	Обозначение
1..12	12 байт максимум, может быть меньше. Это зависит от длины номера SMSC или вообще может быть 00h – тогда номер SMSC берется из настроек телефона (Sim-карты)	SCA	Service Center Address
1	один байт, в котором задается тип PDU – состоит в свою очередь из одно и двухбитовых полей. Именно битами в PDU Type различаются форматы передаваемых и принимаемых сообщений. В принимаемых сообщениях последние два бита этого байта равны 00b, а в передаваемых равны 01b – в нашем случае 00b	PDU Type	Protocol Data Unit Type
2..12	Адрес (номер телефона) отправителя сообщения или, попросте, от кого пришло SMS, телефон отправителя	OA	Originator Address
1	Идентификатор протокола – указывает SMSC, как обрабатывать передаваемое сообщение (как факс, голосовое сообщение и т.д.)	PID	Protocol Identifier

1	Схема кодирования данных в поле UD	DCS	Data Coding Scheme
7	Параметр, который указывает время получения сообщения SCSM	SCTS	Service Centre Time Stamp
1	Длина поля UD	UDL	User Data Length
0..140	140 байт данных пользователя	UD	User Data

6. SCA - Service Center Address - адрес сервисного центра рассылки коротких сообщений

Поле SCA (поле адрес сервисного центра) предназначено для задания номера сервисного центра коротких сообщений. Если задан \$00 - то адрес (номер) SCA берется из Sim-карты телефона и, в этом случае, перед передачей SMS его (SCA) надо прописать в телефон командой `at+cscs="+MMMxxxxxxxx"`.

Внимание: некоторые производители в случае задания \$00 вообще требуют опустить этот байт!!

Если первый байт не \$00 - то это адрес (номер) SC или SMSC (это одно и то же) - номер сервисного центра рассылки коротких сообщений (Service Centre или Short Message Service Centre).

Он (адрес SCA) задается следующим путем (все остальные адресные поля задаются идентично):

Длина в байтах	Пояснение	Поле	Обозначение
1	длина поля SCA	Len	Length
1	тип номера SC: Старший бит всегда 1. Бит Значение Пояснение 7 - всегда 1 6..4 тип номера 000 - неизвестный; 001 - интернациональный; 010 - национальный; 011 - принятый в сети; 100 - тип подписчика сети; 101 - алфавитноцифровой; 110 - сокращенный; 111 - зарезервирован. 3..0 тип набора 0000 - неизвестный; 0001 - ISDN; 0010 - X.121; 0011 - телекс; 1000 - национальный; 1001 - частный (private); 1010 - ERMES; 1111 - зарезервирован.	Type of Number	Type of Number

	<p>Остальные комбинации зарезервированы.</p> <p>Нас интересуют всего два варианта: - \$81 - номер (неизвестный), то есть впереди него стоят символ + и код страны. При кодировании номера символ + отбрасывается, ISDN; - \$91 - номер интернациональный, то есть впереди него стоят символ + и код страны. При кодировании номера символ + отбрасывается, ISDN.</p> <p>На самом деле все остальные биты этого байта также имеют смысл. Для любопытных см. (3. GSM 03.40 version 6.0.0, пп. 9.1.2.5 и GSM 04.08 version 6.7.1, пп. 10.5.4.6) Формулировка, принятая в 1 - «национальный» - неверна, так как на самом деле это «неизвестный». Если пользователь или сеть не знает тип номера («неизвестный»), тогда необходимо писать полный адрес с префиксом. То есть при типе номера \$81 номер надо писать в виде, например, 371 123456, с отброшенным символом +. Установленный младший бит обозначает ISDN/telephone numbering plan</p>		
0..8	Адрес SMSC - номер телефона Центра Обслуживания Коротких Сообщений		

Внимание :

В документе (1) при определении длины номера (SMSC, DA OA) допущена ошибка. Для правильного определения адресных полей следует руководствоваться документом (3) GSM 03.40 version 6.0.0. Именно из него и взято следующее определение:

«The Address-Length field is an integer representation of the number of useful semi-octets within the Address-Value field» - поле Длина Адреса это целое число, которое представляет собой количество полезных полуоктетов в поле Адреса.

По-русски это звучит так: Длина адреса определяется, как количество цифр в номере! В это количество входят только цифры, то есть ни знак + для международного номера, ни дополняющий до четного количества символ F (см. ниже) цифрами НЕ являются и в длину не входят!

Адрес (номер) должен содержать четное количество цифр. Если оно нечетное, то адрес (номер) дополняется символом «метка конца номера», который кодируется как F шестнадцатеричное и в длину адреса не входит!

Принятый при неизвестном типе номера символ + отбрасывается и в PDU не пишется!

Первые 2 байта - длина и тип адреса (номера) - записываются в поле SCA в обычном порядке - то есть слева старшая тетрада, затем младшая. С самим номером необходимо поступать иначе - тетрады в байтах переставляются местами.

Адреса (номера) записываются десятичными цифрами (BCD-кодирование) и в них допускаются только цифры и символ Fh. Если же в поле адреса (номера) встречаются тетрады, не являющиеся BCD цифрами, то они заменяются следующим образом:

код	1010b	на символ	*
	1011		#
	1100		a
	1101		b
	1110		c

(Куда при этом вы позвоните, вам придется выяснять самим!)

Пример 1: номер +37212345678 последовательно превращается:

- 37212345678 - убрали +;
- 37212345678 - определили длину адреса `Length('37212345678') = 11d` (десятичное) или `0Ah` (шестнадцатеричное);
- 37212345678F - добавили в конец F, так длина нечетная;
- 7312325476F8 - переставили тетрады;
- 917312325476F8 - добавили интернациональный тип номера 91h;
- 0A917312325476F8 - добавили длину - и вот вам результат, то есть

номер +37212345678 превратился в 0A917312325476F8

7. Protocol Data Unit Type (PDU type) - тип пакета данных протокола

Поле Pdu Type состоит из одного байта. В общем случае в двоичном виде для принятого сообщения этот байт имеет вид `xxxxxx00b`, а для передаваемого сообщения - `xxxxxx01b`. Биты, обозначенные как x, описаны ниже. Нумерация битов слева направо, старший (самый левый) бит - 7-ой, младший (самый правый) - 0-ой.

7.1. PDU type SMS-Submit (Mobile Originated) - тип PDU для передаваемого SMS

Биты	7	6	5	4..3	2	1..0
Обозначение	RP	UDHI	SRR	VPF	RD	MTI

Бит	Пояснение	Обозначение	Значение
7	Параметр Reply Path 0 - RP не установлен 1 - RP установлен	Rp	Запрос ответа от стороны, принимающей сообщение. Если запрос ответа установлен, то принимающая сторона пытается ответить с тем же SCA для повышения вероятности получения подтверждения передающей стороной. Кроме того, установка этого бита позволяет отвечать устройствам, которые по

			умолчанию не имеют прав на посылку сообщений. Кто платит за ответ определяется оператором мобильной связи.
6	<p>Параметр, определяющий наличие заголовка в UD (данных пользователя)</p> <p>0 - UD содержит только данные</p> <p>1 - UD содержит в добавление к данным и заголовок.</p>	UDHI	<p>Формат заголовка в (1) не описан. В (2) сказано, что большинство мобильных телефонов не поддерживают формат заголовка, который введен для будущего использования со сжатými сообщениями. Нормальное значение этого бита 0. Любопытных отсылают к документу GSM 03.40. Мы не любопытны и устанавливаем этот бит в 0b. Наличие заголовка требует CellBroadcast Message, но такое сообщение может генерировать только оператор мобильной связи.</p>
5	<p>Параметр Status Report Request</p> <p>0 - не требуется</p> <p>1 - требуется</p>	SRR	<p>Запрос на получение отчета. Установленный бит запрашивает от SMSC SMS-Status-Report после получения SMSC SMS_Submit. Этот параметр отличается от RP тем, что RP запрашивает ответ от принимающей стороны, а SRR - от SMSC.</p>
4..3	<p>Параметр Validity Period Format, определяющий формат поля VP</p> <p>4 3</p> <p>0 0 - поле VP отсутствует</p> <p>0 1 - резерв в Siemens, расширенный формат для SonyEricsson</p> <p>1 0 - поле VP использует относительный формат</p> <p>1 1 - поле VP использует абсолютный формат</p>	VPF	<p>Согласно (2) чаще всего достаточно относительного формата представления VP. Следовательно, эти два бита устанавливаем в 10b. Подробнее см. п.п. XXXX</p>
2	<p>Параметр Reject Duplicates говорит SMSC о необходимости удалять одинаковые сообщения, если они еще не переданы. Одинаковыми считаются сообщения с совпадающими MR (Message Reference) и DA (Destination Address) и поступившие от одного OA (Originator Address)</p> <p>0 - не удалять</p> <p>1 - удалять</p>	RD	<p>Согласно (2) обычно этот параметр может быть установлен в 0b</p>
1..0	<p>Message Type Indicator</p> <p>Биты Направление передачи</p>	MTI	<p>Устанавливаем в 01</p>

1 0	MS -> SC	SC -> MS		
0 0	SMS-DELIVER REPORT	SMS-DELIVER		
1 0	SMS-COMMAND	SMS-STATUS REPORT		
0 1	SMS-SUBMIT	SMS-SUBMIT REPORT		
1 1	RESERVED			

7.2. PDU type SMS-Deliver (Mobile Terminated)- тип PDU для принятого SMS

Биты	7	6	5	4..3	2	1..0
Обозначение	RP	UDHI	SRR	не используются	MMS	MTI

Бит	Пояснение	Обозначение	Значение
7	Параметр Reply Path 0 - RP не установлен 1 - RP установлен	Rp	В (2) это поле определяется, как идентичное для PDU SMS_SUBMIT, то есть Запрос ответа.
6	Параметр, определяющий наличие заголовка в UD (данных пользователя) 0 - UD содержит только данные 1 - UD содержит в добавление к данным и заголовок.	UDHI	Формат заголовка в (1) не описан. В (2) сказано, что большинство мобильных телефонов не поддерживают формат заголовка, который введен для будущего использования со сжатыми сообщениями. Нормальное значение этого бита 0b. Любопытных отсылают к документу GSM 03.40. Заголовок предназначен для формирования «длинных» сообщений, собираемых из нескольких коротких сообщений.
5	Параметр Status Report Indication 0 - отчет не будет передан ME 1 - отчет будет передан	SRI	Устанавливается на стороне SMSC.
4..3	В этом типе PDU не используются		Не анализируются на принимающей стороне.
2	Параметр More Message to Send 0 - Ожидаются еще сообщения на стороне SMSC для MS - скорее всего это ожидается продолжение длинного сообщения 1 - Сообщения не ожидаются	MMS	Устанавливается на стороне SMSC.
1..0	Message Type Indicator Биты Направление передачи 1 0 SC -> MS MS -> SC	MTI	Установлено в 00b

0 0	SMS-DELIVER	SMS-DELIVER REPORT	
1 0	SMS-STATUS REPORT	SMS-COMMAND	
0 1	SMS-SUBMIT REPORT	SMS-SUBMIT	
1 1	RESERVED		

7.3. Message Reference

Поле MR предназначено для хранения количества переданных SMS. Это число представлено, как целое, занимает один байт и может иметь значения от 0 до 255. Определено только для передаваемых сообщений и генерируется MS. Тем не менее вы должны установить это значение в 00h.

7.4. Originator Address OA и Destination Address DA – адрес источника (создателя) и адрес приемника (получателя)

Поля OA (адрес источника) и DA (адрес приемника) представляют собой адрес отправителя SMS и адрес получателя SMS, то есть, проще говоря, номера телефонов отправителя и получателя SMS. Формируются эти поля идентично полю SCA- Service Center Address – и состоят из байта длины, байта типа номера и собственно номера.

7.5. Protocol Identifier PID – идентификатор протокола

Поле PID (идентификатор протокола) предназначено для того, чтобы транспортный протокол мог обмениваться информацией с протоколом высшего уровня. PID описывает высокоуровневые протоколы или взаимодействующие устройства, от которых или для которых предназначена SMS. SMSC (Центр Обслуживания Коротких Сообщений) сети может обеспечивать преобразования SMS в e_mail, сообщения оповещения, формат Телекса и т.д. и наоборот. Однако в нашем случае значения этого поля сообщают MS (приемнику сообщения) о методе замены коротких сообщений. Если значение этого поля 00h, то короткие сообщения не заменяются, а если значение этого поля лежит в пределах 41h..47h, то происходит замена короткого сообщения, имеющего одинаковые PID, SCA и OA на новое сообщение. Если нет сообщения, которое надо заменить, сохранение нового сообщения происходит нормальным методом. **Обратите внимание на то, что это происходит на стороне MS!**

7.6. Data Coding Scheme DCS – схема кодирования данных

Поле DCS (схема кодирования данных) представляют собой схему кодирования поля UD, то есть поля данных. Схема кодирования данных подробно специфицирована в документе (6) GSM 03.38 V7.0.0 (1998-07).

Поле DCS представляет собой байт из двух тетрад по 4 бита. Старшая тетрада (биты с 7 по 4) определяют группу кодирования, а младшая тетрада (биты с 3 по 0) определяют специфические данные для группы кодирования:

Группа кодирования, биты 7..4	
Биты 7 6 5 4 0 0 x x	<p>Общее кодирование данных (General Data Coding)</p> <p>Старшие два бита всегда 00b. Бит 5 определяет сжатие текста. Бит 4 определяет формирование класса сообщения.</p> <p>Бит 5 0 – текст не сжат; 1 – текст сжат, согласно алгоритма сжатия GSM (для любопытных см. GSM 03.42).</p> <p>Бит 4 0 – биты 1 и 0 в младшей тетраде зарезервированы и не обозначают класс сообщения; 1 – биты 1 и 0 в младшей тетраде определяют класс короткого сообщения следующим образом:</p> <p>Биты 1 и 0 – Message Class – класс сообщения 0 0 – Класс 0; 0 1 – Класс 1 – специфический для Mobile Equipment; 1 0 – Класс 2 – специфический для SIM specific message; 1 1 – Класс 3 – специфический для Terminal Equipment (для любопытных см GSM 07.05) (Коротко классы сообщений будут рассмотрены ниже).</p> <p>Биты 3 и 2 определяют алфавит сообщения следующим образом: 0 0 Алфавит по умолчанию; 0 1 8-ми битовые данные; 1 0 UCS2 (16 битовые данные) – Unicode; 1 1 Зарезервировано.</p> <p>Внимание : Особый случай, когда все биты DCS равны 0, то есть DCS = 00h говорит о применении Default GSM Alphabet (алфавита GSM по умолчанию) – именно такой вариант рекомендует применять Siemens.</p>
Биты 7 6 5 4 Значения с 0 1 0 0 по 1 0 1 1	Зарезервировано
Биты 7 6 5 4 1 1 0 0	<p>Отмена сообщений (Discard Message)</p> <p>Биты с 3 по 0 кодируются точно так же, как в группе 1101. Такая группа кодировки позволяет ME (то есть телефону) отказаться от сохранения содержимого сообщения и отобразить текст сообщения только на экране (индикация).</p>
Биты 7 6 5 4 1 1 0 1	<p>Сохранение сообщений (Store Message)</p> <p>Эта группа позволяет выводить индикацию о состоянии сообщений, которые ожидаются системой. Телефон может отображать это состояние иконкой на экране или каким-либо другим методом. Текст сообщения кодируется всегда</p>

	<p>Default GSM Alphabet (алфавитом GSM по умолчанию). Когда сообщение поступает с такой группой кодировки телефон сохранит текст сообщения и отобразит текст сообщения на экране (индикация).</p> <p>Бит 3 имеет следующие значения 0 - выключение индикации; 1 - включение индикации.</p> <p>Бит 2 зарезервирован и установлен в 0.</p> <p>Биты 1 и 0 определяют тип индикации: 0 0 - ожидается сообщение голосовой почты (Voicemail Message Waiting); 0 1 - ожидается факс (Fax Message Waiting); 1 0 - ожидается электронное письмо (E-Mail Message Waiting); 1 1 - ожидается «другое» сообщение (Other Message Waiting).</p> <p>«Другое» сообщение может отображаться каким-либо способом по усмотрению производителя ME. На настоящий момент не стандартизовано, не рекомендуется к применению.</p>
<p>Биты 7 6 5 4 1 1 1 0</p>	<p>Сохранение сообщений (Store Message UCS2) Биты с 3 по 0 кодируются точно так же, как в группе 1101. Единственное отличие в том, что текст сообщения кодируется всегда несжатым UCS2 алфавитом.</p>
<p>Биты 7 6 5 4 1 1 1 1</p>	<p>Кодирование данных/Класс сообщения (Data coding/message class) Бит 3 зарезервирован и установлен в 0. Бит 2 Кодирование сообщения (Message coding): 0 - алфавит по умолчанию (Default alphabet); 1 - 8-битовое кодирование.</p> <p>Биты 1 и 0 - Message Class - класс сообщения 0 0 - Класс 0; 0 1 - Класс 1 - специфический для Mobile Equipment; 1 0 - Класс 2 - специфический для SIM specific message; 1 1 - Класс 3 - специфический для Terminal Equipment (для любопытных см GSM 07.05) (Коротко классы сообщений будут рассмотрены ниже).</p>

Алфавит по умолчанию (Default alphabet) говорит о том, что UD (данные пользователя) кодируются 7-битовым алфавитом, при этом восемь символов запаковываются в семь байтов и сообщение может содержать до 160 символов.

При 8-ми битовом кодировании применяется INTEL ASCII-HEX таблица и сообщение в таком случае не может содержать более 140 символов.

Внимание: Siemens рекомендует использовать сообщения класса 2 (Class 2 message - SIM specific) или группу кодировки 0000b (так как при этом текст сообщения не сжат и класс сообщения не определен). (Полученное сообщение будет обработано самим телефоном и результат этой обработки будет зависеть от производителя и

модели телефона! Впрочем, и во всех остальных случаях правильность отображения и сохранения сообщения будет зависеть от производителя и модели телефона, если вы сами не наделаете ошибок!)

7.7. Message Class - Классы сообщений

Classless SMS - сообщение без класса

Обычно сообщение, посланное мобильным телефоном. Оно будет сохранено в доступной памяти, обычно в памяти самого телефона.

Class 0 SMS - сообщение класса 0

Сообщение класса 0 нигде не сохраняется и выводится напрямую на дисплей телефона. В случае, если у мобильного модуля нет дисплея имеется возможность перенаправить сообщение напрямую терминальному оборудованию (TE) при помощи AT-команды AT+CNMI=3,2. Siemens называет такой вариант получения сообщений «окружным путем» (roundabout way).

Class 1 SMS - сообщение класса 1

Сообщение класса 1 сохраняется в памяти ME (мобильного оборудования), если такая память существует и доступна. Иначе сообщение сохраняется в памяти SIM-карты.

Class 2 SMS - сообщение класса 2

Сообщение класса 2 сохраняется в памяти SIM-карты.

Class 3 SMS - сообщение класса 3

Сообщение класса 3 обычно передается терминальному оборудованию или приложению 3. Этот процесс контролируется и управляется AT-командой AT+CNMI. (Для любопытных - см. GSM specification 03.38.)

Внимание: При посылке сообщения в 8-ми битовой кодировке на обычный мобильный телефон результаты могут быть совершенно непредсказуемы и полученное сообщение может быть совершенно искажено. Это возникает вследствие того, что обычно телефоны понимают только GSM кодировку по умолчанию, а она 7-ми битовая. При распаковке UD телефоном и происходят такие искажения. Sony-Ericsson рекомендуют попробовать и посмотреть, что получится. Нынешние мобильные телефоны, скорее всего, этого «качества» лишены и понимают все кодировки, но рекомендуется все это проверять на используемых вами моделях телефонов и GSM-модемов.

7.8. Service Center Time Stamp - SCTS - метка времени Short Message Service Center (центра обслуживания коротких сообщений)

Поле SCTS предназначена для того, чтобы сообщить принимающему устройству время получения короткого сообщения транспортным протоколом SMSC. Метка времени состоит из 7 байтов и включена в любое SMS-DELIVER, которое было доставлено в SMSC и отображает локальное время следующим образом:

Номер байта	1 байт	2 байт	3 байт	4 байт	5 байт	6 байт	7 байт
Обозначение	год	месяц	день	Час	Минута	Секунда	Пояс
Порядок тетрад	2 1	2 1	2 1	2 1	2 1	2 1	2 1

Часовой пояс указывает различие, выраженное в четвертях часа, между местным временем и временем по Гринвичу (GMT - Greenwich Main Time).

Внимание: Строка «Порядок тетрад» говорит о том, что в каждом байте тетрады

переставлены!

Пример: 2008 февраль 23 17 часов 32 минуты 16 секунд в Гринвиче было получено SMS, тогда метка времени должна иметь следующий вид (выполняем последовательно)

```
2008 февраль 23 17 часов 32 минуты 16 секунд Мы в Гринвиче
 08 02 23 17 32 16 00 - записываем в символьном виде;
 80 20 32 71 23 61 00 - переставляем тетрады и
```

80203271236100 - результат.

7.9. Validity Period VP - Время жизни сообщения

Время жизни сообщения - указывает SMSC время, по истечении которого сообщение недействительно (valid - законный), проще говоря - это время, по истечении которого сообщение в памяти SMSC уничтожается, если SMSC не получил уведомление о доставке сообщения получателем.

Формат VP определяется полем VPF (Validity Period Format).

Значение VPF	Формат VP	Длина поля VP в байтах
00b	поле VP отсутствует	0
01b	резерв в Siemens, расширенный формат для SonyEricsson	-
10b	поле VP использует относительный формат	1
11b	поле VP использует абсолютный формат	7

При задании относительного формата VP кодируется одним байтом в соответствии со следующей таблицей:

Значение VP, десятичное	Как считать VP
0 .. 143	$(VP + 1) * 5$ минут = максимально можно задать 12 часов с шагом 5 минут
144 .. 167	12 часов + $(VP - 143) * 30$ минут = максимально можно задать 24 часа
168 .. 196	$(VP - 166) * 1$ день = максимально можно задать 30 дней
197 .. 255	$(VP - 192) * 1$ неделю = максимально можно задать 63 недели

При задании абсолютного формата VP кодируется аналогично SCTS и представляет собой время, по истечении которого SMSC может удалять сообщение.

7.10. User Data Length UDL - длина данных пользователя

Длина данных пользователя UDL определяет количество данных в User Data (данных пользователя). UDL определяет количество данных двумя разными способами, в зависимости от схемы кодирования данных DCS. Если определен алфавит по умолчанию, то каждый передаваемый символ представляет собой 7 бит (септет) и тогда количество данных определяют не байты, а символы. То есть, если мы передаем восемь 7-ми бит кодированных символов (это будет $8 * 7 = 56$ бит) длина будет 8, хотя сообщение займет 7 байтов = $56 / 8$.

Если же мы применяем 8-ми или 16-ти битовую кодировку, то UDL равно количеству байтов.

7.11. User Data UD - данные пользователя

Данные пользователя - это собственно данные, которые передаются в сообщении. Их общая длина в байтах не может превышать 140 байтов. В случае применения 8-ми или USC2 кодировки данные никак не упаковываются и, соответственно, на принимающей стороне никак не распаковываются. В случае применения кодировки Default GSM alphabet - 7-ми битовой кодировки, данные упаковываются и в 140 байтах передаются 160 символов 7-ми битовой кодировки. На принимающей стороне, соответственно, эти данные должны быть распакованы.

7.12. Метод упаковки данных пользователя

При 7-ми битовой кодировке данные упаковываются в байты следующим образом, описанным в документе (6) GSM 03.38 version 7.0.0 Release 1998 GSM 03.38 V7.0.0 (1998-07) (Phase 2+):

- 1 байт - старший бит это младший бит 2-го символа + 7 бит первого символа
- 2 байт - 2 старших бита это младшие два бита 3-го символа + 6 оставшихся бит 2-го символа
- 3 байт - 3 старших бита это младшие биты 4-го символа + 5 оставшихся бит 3-го символа и так далее.

Поясним это на примере.

Например, мы хотим передать восемь первых строчных букв (символов) латинского алфавита - abcdefgh. Каждый символ кодируется 7 битами, обозначим их цифрами, где 0 - самый младший (правый) бит, а 6 - самый старший (левый) бит. В результате у нас получилась следующая таблица:

Биты Символ	6	5	4	3	2	1	0
a = 61h	a6	a5	a4	a3	a2	a1	a0
b = 62h	b6	b5	b4	b3	b2	b1	b0
c = 63h	c6	c5	c4	c3	c2	c1	c0
d = 64h	d6	d5	d4	d3	d2	d1	d0
e = 65h	e6	e5	e4	e3	e2	e1	e0
f = 66h	f6	f5	f4	f3	f2	f1	f0
g = 67h	g6	g5	g4	g3	g2	g1	g0
h = 68h	h6	h5	h4	h3	h2	h1	h0

Теперь упакуем эти данные в байты, согласно правилам, описанным выше и получим

следующую таблицу:

Биты Байты	7	6	5	4	3	2	1	0
1	b0	a6	a5	a4	a3	a2	a1	a0
2	c1	c0	b6	b5	b4	b3	b2	b1
3	d2	d1	d0	c6	c5	c4	c3	c2
4	e3	e2	e1	e0	d6	d5	d4	d3
5	f4	f3	f2	f1	f0	e6	e5	e4
6	g5	g4	g3	g2	g1	g0	f6	f5
7	h6	h5	h4	h3	h2	h1	h0	g6

Как не трудно заметить восемь семибитовых символов упаковались в семь байтов.

Теперь сделаем упаковку с реальными кодами символов abcdefgh:

Биты Символ	6	5	4	3	2	1	0
a = 61h	1	1	0	0	0	0	1
b = 62h	1	1	0	0	0	1	0
c = 63h	1	1	0	0	0	1	1
d = 64h	1	1	0	0	1	0	0
e = 65h	1	1	0	0	1	0	1
f = 66h	1	1	0	0	1	1	0
g = 67h	1	1	0	0	1	1	1
h = 68h	1	1	0	1	0	0	0

Упаковываем в байты

Биты Байты	7	6	5	4	3	2	1	0	Hex
1	0	1	1	0	0	0	0	1	61
2	1	1	1	1	0	0	0	1	F1
3	1	0	0	1	1	0	0	0	98
4	0	1	0	1	1	1	0	0	5C
5	0	0	1	1	0	1	1	0	36
6	1	0	0	1	1	1	1	1	9F
7	1	1	0	1	0	0	0	1	D1

В результате строка символов abcdefgh превратилась в строку символов 61F1985C369FD1. Именно эта строка и должна быть передана в ME в поле User Data. То есть каждый байт данных при передаче в ME кодируется двумя шестнадцатеричными символами.

При 8-ми битовой или USC2 кодировке никакая упаковка не производится и данные записываются в UD «как есть», только каждый символ заменяется на свой код, записанный в шестнадцатеричной форме.

Например, при кодировке USC2 строка «Unicode – это наше все» превращается в

0055006E00690063006F00640065002003E60020044D0442043E0020043D0430044804350020043204410435.

Здесь следует отметить тот факт, что сейчас идет речь о кодировании, принятом в сетях GSM. Эти системы кодирования не обязательно совпадают с системами кодирования, принятыми в применяемых вами операционных системах или процессорах. Так что впереди вас ждет еще много увлекательного. Я, например, с удивлением узнал, что кодировка (правильно – кодовая страница) Win1251 отличается от кодировки ISO 8859-5, UTF8 это не совсем UTF16, а KOI-8 это практически стандарт русских Unix- систем.

Что происходит при неправильной раскодировке замечательно описано в (10).

Статьи (11) и (12) помогут вам войти в жуткий (другого слова нет) мир перекодировок символов.

8. Заключение

При разборе формата PDU складывается впечатление, что в этот формат пытались засунуть все возможные виды представления данных. Это приводит к потере прозрачности и заметно усложняет понимание формата. Но делать нечего, PDU теперь долгое время будет переходить из стандарта в стандарт мобильной связи. Он (формат) уже есть в sms, будет и в mms. И мы еще долгое время будем ковыряться в этих битах и байтах, символах и кодировках, рассчитывая найти все свои и чужие ошибки. Удачи нам!

Рига, февраль 2008

Алексей Печерский
(AlexPP)

P.S. Нашел в Сети:

03.06.2007 в 13:02:36 на <http://lenta.parttimer.ru/1114/> – «нужна работающая на СИ процедура преобразования PDU-текста, получаемого по ат-команде "AT+CMGR=", в обычный текст ASCII с поддержкой русских sms. Пример – TPduToMsg(char encodedmsg[],char pdu_msg[]). передаем pdu_msg[], получаем e-mail, под CodeVisionAVR? на СИ.»

Бюджет проекта – 4 (!) доллара, то есть, по нашим ценам (Латвия), пачка сигарет и бутылка пива!

АП

Приложение 1.

Таблица символов алфавита GSM по умолчанию (GSM default alphabet):

Биты				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	Δ	SP	0	i	P	ı	p
0	0	0	1	1	£	_	!	1	A	Q	a	q
0	0	1	0	2	§	Ф	"	2	B	R	b	r
0	0	1	1	3	¥	Г	#	3	C	S	c	s
0	1	0	0	4	è	Λ	α	4	D	T	d	t
0	1	0	1	5	é	Ω	%	5	E	U	e	u
0	1	1	0	6	ù	Π	&	6	F	V	f	v
0	1	1	1	7	ì	Ψ	'	7	G	W	g	w
1	0	0	0	8	ò	Σ	(8	H	X	h	x
1	0	0	1	9	Ç	Θ)	9	I	Y	i	y
1	0	1	0	A	LF	Ε	*	:	J	Z	j	z
1	0	1	1	B	∅		+	;	K	Д	k	д
1	1	0	0	C	ø	Æ	,	<	L	Ö	l	ö
1	1	0	1	D	CR	æ	-	=	M	Ñ	m	ñ
1	1	1	0	E	Å	β	.	>	N	Ü	n	ü
1	1	1	1	F	å	É	/	?	O	Ş	o	à

Некоторые символы в таблице могут быть изменены. Например 0x1B зарезервирован для знака Евро.

Приложение 2.

Таблица символов UCS2-C0 controls and Basic Latin Range 0000-007F

(взято из документа C0 controls and Basic Latin Range 0000-007F by Unicode Inc., <http://www.unicode.org/charts/PDF/U0000.pdf>)

Биты				b7	0	0	0	0	0	0	0	0	0
				b6	0	0	0	0	1	1	1	1	
				b5	0	0	1	1	0	0	1	1	
				b4	0	1	0	1	0	1	0	1	
b3	b2	b1	b0		0	1	2	3	4	5	6	7	
0	0	0	0	0	Nul	DLE	SP	0	@	P	'	p	
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q	
0	0	1	0	2	STX	DC2	''	2	B	R	b	r	
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s	
0	1	0	0	4	EOT	DC4	\$	4	D	T	d	t	
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u	
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v	
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w	
1	0	0	0	8	BS	CAN	(8	H	X	h	x	
1	0	0	1	9	HT	EM)	9	I	Y	i	y	
1	0	1	0	A	LF	SUB	*	:	J	Z	j	z	
1	0	1	1	B	VT	ESC	+	;	K	[k	{	
1	1	0	0	C	FF	FS	,	<	L	\	l		
1	1	0	1	D	CR	GS	-	=	M]	m	}	
1	1	1	0	E	SO	RS	.	>	N	^	n	~	
1	1	1	1	F	SI	US	/	?	O	_	o	DEL	

Приложение 3.

Таблица символов UCS2 Cyrillic Range 0400-04FF (взято из статьи в Википедии «Кириллица в Юникоде» и документа Cyrillic Range 0400-04FF by Unicode Inc., <http://www.unicode.org/charts/PDF/U0400.pdf>)

В Unicode для кириллицы выделено два раздела:
Cyrillic – кириллица (U+0400 – U+04FF)
Cyrillic Supplement – кириллица (приложение) (U+0500 – U+052F).

Символы кириллицы можно разделить на 3 группы:

1. U+0400 – U+045F – это символы из ISO 8859-5, но перемещённые вверх на 864 позиции (0360h); (**Внимание: для кодировки Win1251 – перемещенные на 0350h позиций вверх!**).
2. U+0460 – U+0489 – исторические буквы славянских алфавитов;
3. U+048A – U+052F – это дополнительные буквы для разных языков, использующих кириллицу.

Нас интересуют в первую очередь группа U+0400 – U+045F – в ней, начиная с кода 0410 по код 044f включительно расположены буквы традиционного современного русского алфавита. Исключения составляют буквы Ё и ё, которые имеют коды 0401 и 0451 соответственно. (Для любопытных – см. документ Cyrillic Range 0400-04FF by Unicode Inc. – там, в частности, приведена таблица всех символов кириллицы, которые нашли свое место в Unicode. Среди них особенно хороши символы «густое придыхание» и «тонкое придыхание» !)

Биты				b7	0	0	0	0
				b6	0	0	0	1
				b5	0	1	1	0
				b4	1	0	1	0
b3	b2	b1	b0		1	2	3	4
0	0	0	0	0	А	Р	а	р
0	0	0	1	1	Б	С	б	с
0	0	1	0	2	В	Т	в	т
0	0	1	1	3	Г	У	г	у
0	1	0	0	4	Д	Ф	д	ф
0	1	0	1	5	Е	Х	е	х
0	1	1	0	6	Ж	Ц	ж	ц
0	1	1	1	7	З	Ч	з	ч
1	0	0	0	8	И	Ш	и	ш
1	0	0	1	9	Й	Щ	й	щ
1	0	1	0	А	К	Ъ	к	ъ
1	0	1	1	В	Л	Ы	л	ы
1	1	0	0	С	М	Ь	м	ь
1	1	0	1	Д	Н	Э	н	э
1	1	1	0	Е	О	Ю	о	ю
1	1	1	1	Ф	П	Я	п	я

Список литературы:

1. SMS with the PDU mode by Siemens;
2. Application Note Construction of SMS PDU's by Sony Ericsson;
3. GSM 03.40 version 6.0.0 Draft EN (GSM 03.40) V6.0.0 (1998-03);
4. GSM 04.08 version 6.7.1 Release 1997 ETSI EN 300 940 V6.7.1 (2000-06);
5. GSM 07.05 version 5.5.0: January 1998;
6. GSM 03.38 version 7.0.0 Release 1998 GSM 03.38 V7.0.0 (1998-07) (Phase 2+);
7. <http://old.electronix.ru/standatds/sms-pdu/index.php> - SMS and the PDU format;
8. Отправка коротких сообщений (SMS) через сотовый телефон, подключенный к компьютеру
http://www.ixbt.com/mobile/review/comp_sms.shtml, Олег Ямщиков;
9. <http://www.unicode.org/charts/PDF/U0000.pdf> C0 controls and Basic Latin Range 0000-007F by Unicode Inc.;
10. <http://www.unicode.org/charts/PDF/U0400.pdf> Cyrillic Range 0400-04FF by Unicode Inc.;
11. <http://ru.wikipedia.org/wiki/%D0%9A%D1%80%D0%BE%D0%BA%D0%BE%D0%B7%D1%8F%D0%B1%D1%80%D1%8B> статья в Wikipedia, посвященная преобразованию текста и получению «крокозябр» - очень познавательно!;
12. http://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D0%B4%D0%BE%D0%B2%D0%B0%D1%8F_%D1%81%D1%82%D1%80%D0%B0%D0%BD%D0%B8%D1%86%D0%B0 - статья «Кодовая страница»;
13. <http://ru.wikipedia.org/wiki/Windows-1251> - статья «Win 1251».