CAPACITY ANALYSIS OF THE GSM SHORT MESSAGE SERVICE

Nilesh Agarwal   Leena Chandran-Wadia   Varsha Apte

Indian Institute of Technology Bombay
Powai, Mumbai 400 076, India

ABSTRACT

We present a simple queueing model to analyze the SMS capacity of a GSM cellular network. The analysis is based on a detailed study of GSM channels and their usage by SMS and other traffic. We saw that SMS is sent on the same control channels that are used by voice call set-up and location updates. Our results show that if SMS volumes continue to increase dramatically, the control channels used by SMS can become a bottleneck, thus increasing the blocking probability of SMS as well as voice calls.

I. INTRODUCTION

The GSM mobile telephony system provides text messaging through an extremely popular service called the *Short Message Service (SMS)*. The number of GSM users worldwide has exceeded 900 million and the number of SMS messages exchanged each month is approximately 15 billion. In India SMS usage is increasing rapidly as more and more innovative uses are being found, such as contests, shopping and location based services. Apart from good volumes on normal days, SMS usage in India and abroad reaches abnormally high peaks on special days such as the New Year and other popular festivals.

A natural question that arises when one sees such phenomenal usage of a service is: What happens to the performance of this service, as volumes increase beyond expectations? Or, how are *other* services that are provided on the same network affected, with increasing SMS usage? The work described here tries to answer these questions.

The main contributions of this paper are: 1) A comprehensive description of the SMS within GSM, including details of GSM channel assignment, which is the key to understanding SMS capacity of a GSM network, 2) Identification of traffic that contends with SMS for the same resources, and 3) Running some what-if scenarios using the Erlang blocking model for SMS capacity.

In the following, Section 2 provides an overview of the short message service. Section 3 provides details of GSM channel assignment for SMS and other traffic. Section 4 describes the model and Section 5 some results. We conclude the paper in Section 6.

II. SERVICE DESCRIPTION

SMS offers two types of service: One is the cell broadcast service in which a message is sent to all Mobile Subscribers (MS) in the cell that are subscribed to the particular service. The other is a point-to-point (PTP) service in which individual mobile subscribers send and receive messages of limited size from their GSM mobiles. In this work we are concerned with the latter. The point-to-point short message service can be either Mobile originated (MO-SM) or Mobile Terminated (MT-SM). Mobile originated short messages are transported from an MS to the short message Service Center (SC). MT-SMs are transported from the SC to the handset.
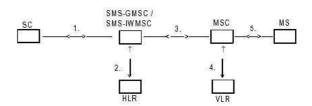

Fig. 1. **Network Elements and Architecture**

The network elements involved in the transfer of short messages are shown in Figure 1. The figure shows the sequence of steps involved in delivering a MT-SM to an MS. A MT-SM will start from the SC, then will be forwarded to a Gateway MSC (Mobile Switching

Centre), which will locate the MS by querying the Home Location Register (HLR), forward it to the appropriate MSC, which in turn queries the Visitor Location Register (VLR) for subscriber information, and then delivers the message to the MS. The final delivery of the message is through GSM channels, the capacity of which is the focus of this paper, and described next.

III. GSM LOGICAL AND PHYSICAL CHANNELS

GSM uses a large number of logical channels which are mapped onto the physical channels in the radio path [1]. The logical channels consist of two groups: the traffic channels (TCH) and the control channels. The latter consist of broadcast control (BCCH), common control (CCCH) and dedicated control (DCCH) channels. SMS messages are carried in the dedicated control channels of the network.

The BCCH is a one-to-many unidirectional control channel for broadcasting information from the Base Station (BS) to the MSs. The CCCH is a point-to-multipoint bidirectional control channel primarily intended to carry signaling information necessary for access management. It includes (1) a downlink *Paging Channel* (PCH) for paging MSs (2) an uplink *Random Access Channel* (RACH) used for requesting assignment of a DCCH and (3) a downlink *Access Grant channel* (AGCH) used to assign an MS to a specific DCCH.

A DCCH is a point-to-point, directional control channel which consists of a *stand alone dedicated control channel* (SDCCH) and the *associated control channel* (ACCH). The SDCCH is used before the MS is assigned a TCH. It is used to provide authentication to the MS, for voice call setup and location updates, and assignments of TCHs. It is also used for system signaling during idle periods. A *slow associated control channel* (SACCH) can be linked to either a TCH or an SDCCH. This is a continuous data channel carrying information from the mobile, such as measurement reports of received signal strength.

III-A. Usage of GSM Channels by Different Services

The GSM logical channels are used in different ways by different services. We describe below only those services, that require channels that are also used by the SMS.

1. Voice Call Setup - Voice calls are either mobile originated (MO) or mobile terminated (MT). For MO calls, the MS sends a signaling channel request to the network through the RACH. If a SDCCH is available, the BSC (Base Station Controller) informs the MS of the allocated signaling channel (SDCCH) through the AGCH. Figure 2 shows the complete sequence of control messages exchanged on various logical channels during the setup of a MO voice call [2] (in the figure, BTS denotes Base Transceiver Station). In case of MT calls, the request for a signaling channel is the result of a paging request by the BS on the PCH. The rest of the control messages follow a similar pattern.

Steps	Channel	MS	BTS	BSC	MSC
1. Channel Request	RACH	—▶	—▶		
2. Channel Assigned	AGCH	◀—	◀—		
3. Call Establishment Request	SDCCH	—▶	—▶	—▶	
4. Authentication Request	SDCCH	◀—	◀—	◀—	
5. Authentication Response	SDCCH	—▶	—▶	—▶	
6. Ciphering Command	SDCCH	◀—	◀—	◀—	
7. Ciphering Ready	SDCCH	—▶	—▶	—▶	
8. Send Destination Address	SDCCH	—▶	—▶	—▶	
9. Routing Response	SDCCH	◀—	◀—	◀—	
10. Assign Traffic Channel	SDCCH	—▶	—▶		
11. Traffic Channel Establishment	FACCH	◀—	◀—		
12. Available/Busy Signal	FACCH	◀—			
13. Call Accepted	FACCH	◀—	◀—	◀—	
14. Connection Established	FACCH	—▶	—▶	—▶	
15. Information Exchange	TCH	◀—			▶

Fig. 2. **Mobile Originated Voice Call Setup**

2. Location Update - The location information of each mobile is stored in the HLR and VLR at the MSC. Whenever a mobile is switched on or when it moves from one location area to another, it needs to send that information to the MSC so that the MS can be contacted whenever needed. The sequence of messages exchanged are the same as in Figure 2 except that no traffic channel is assgined and the location update continues on the SDCCH after authentication, ciphering and routing.

3. *SMS* - SMS messages are carried on either SDCCH or SACCH [3] depending on the use of the traffic channel. When the TCH is not allocated, i.e., no voice call or data transfer in progress, the short message is carried on the SDCCH.

If a TCH is allocated before a short message transfer starts, the short message uses the SACCH associated with the TCH. If a TCH is allocated during a short message transfer on an SDCCH (i.e., a voice call or data transfer starts during short message transfer), the

short message transaction stops and continues on the SACCH associated with the TCH. If the voice call or data transfer ends during short message transfer, the short message may either continue on the SACCH associated with the TCH or may stop and continue on a SDCCH.

Once again, the sequence of messages exchanged is similar to that of the voice call setup in Figure 2. In the case of SMS message transfer when a traffic channel is not allocated, the SDCCH is not released after authentication, ciphering and routing but is retained for sending or receiving the message. The time for which the SDCCH channel is held, is calculated in Section 4.

III-B. Channel Allocation in a Cell

Each operator in a cellular area is given a part of the total available frequency bands. The operator divides these allocated frequency bands between the cells in the cluster. Thus, each cell has n (duplex) carriers, named $C0$, $C1$,..., Cn and each carrier has 8 timeslots numbered $TN0$, $TN1$,..., $TN7$. Each combination of a frequency band number and a timeslot number in that band represents a physical channel. Of the total physical channels in a cell, a few are reserved for control channels and the remaining are used for sending and receiving user data i.e., as traffic channels.

Channel $(C0, TN0)$ in a cell is always used for control information. The BCHs are mapped on to this channel, sometimes in combination with other channels. Apart from this control channel, there can be at most three others on which CCCHs are mapped (TN2, TN4, TN6). This number is decided by the operator.

The remaining control channels have SDCCH and associated SACCH. These are not restricted to being allocated either on one of the remaining timeslots on C0 or on TN0 in one of the other frequency bands. Each such channel can serve 8 users (Figure 3).

The GSM logical channels can be mapped onto a physical channel according to some permitted rules [4]. The dedicated control channels SDCCH can be mapped onto physical channels in the following ways.

• FCCH + SCH + BCCH + CCCH + SDCCH/4(0...3) + SACCH/C4(0...3)

• SDCCH/8(0 ..7) + SACCH/C8(0 .. 7)

where SACCH/C is the SACCH associated with SDCCH and the numbers appearing in parenthesis after channel designations are sub-channel numbers.
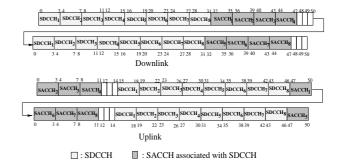

Fig. 3. **Time Organization of SDCCH and associated SACCH**

There are 51 frames in all (one timeslot TN0 in each frame) forming a control multiframe which lasts for 235.5 ms. 26 control multiframes make up a superframe and 2048 superframes constitute a hyperframe.

The SDCCH occupy 32 slots in the 51 TDMA frame control multiframe in eight groups of 4 slots each, each group serving a different user (Figure 3). We focus on SMS messages that are sent on these SDCCH.

In the following section we model SDCCH as the bottleneck resource for which multiple types of GSM traffic contend. As we shall see, the detailed understanding of control channels and their realization in terms of specific timeslots in control multiframes will be used to derive model parameters such as channel holding times.

IV. SMS CAPACITY MODEL

In the previous section we described the sources of traffic that SMS messages contend with namely, voice call setup messages and location updates. We now create a simple queueing model to find the primary measure of interest for a *loss system* like this one - namely, the *blocking probability*. This is the probability that an arriving message will find all channels that it can use, busy, and is an important quality of service measure for SMS and other services that use SDCCH, including voice calls.

IV-A. Model Description

We construct an *Erlang Loss System* ([5]) to model the network since blocked messages are lost, not delayed, in a cellular network. The model is for a single cell in a GSM system and a single operator in the cell.

We model only the control channels which have SD-

CCH mapped onto them. As discussed earlier, each physical channel can have either eight or four SDCCH subchannels, depending upon the organization of control channels. Each of these subchannels serves a different user. In practice, GSM systems do not necessarily allocate SDCCH statically, so the number of "servers" in our queueing model will be variable. We denote this number by s.

Let λ_{sms}, λ_l and λ_v be the arrival rates for SMS messages, location updation and voice call setup respectively. The arrival processes of all the three types of requests are assumed to be Poisson([5]). The maximum size of a single SMS message can not exceed 160 characters (140 octets). (This restriction comes from the MAP protocol of GSM [6].) A message of larger size is broken into fragments of size not exceeding 160 characters. We assume that each of these fragments is treated as a separate SMS message by the system; i.e. after sending one fragment, the SDCCH channel is released. Clearly, a message with size greater then 160 characters is considered to be sent successfully only if all the fragments into which it has been broken have been sent successfully. The effective arrival rate used in the Erlang loss system is then equal to the actual message arrival rate multiplied by the average number of fragments that a message is broken into.

We would like to note here some assumptions made for building the model. First, we assume that the SDCCH is the bottleneck resource, and model only those, although the CCCHs are also needed for sending an SMS message. We also ignore SMS messages that are sent or received while a voice call is in progress, since they use a different channel (SACCH associated with the TCH). Second, we do not model TCH channels explicitly, although blocking a TCH channel will result in re-attempts and therefore, a higher volume of call-setup requests, which will affect our model. Third, SMS message arrivals are assumed to be a Poisson, even though larger messages are sent in fragments and such bulk arrivals are not modeled accurately by the Poisson assumption.

IV-B. Blocking Probability

We model the SDCCH channels as an $M/G/s$ queue with a single "aggregated" customer type. The arrival rate of this aggregate traffic that uses SDCCH channels is given by: $\lambda_c = \lambda_{sms} + \lambda_l + \lambda_v$.

Let the mean service time (i.e. channel holding time) for a single SMS message be μ_{sms}^{-1}, and that of location updation and voice call setup messages be μ_l^{-1} and μ_v^{-1} for respectively.

The unconditional expected service time of an arriving message is then given by $\mu_c^{-1} = \frac{\lambda_{sms}}{\lambda_c} \times \mu^{-1} + \frac{\lambda_l}{\lambda_c} \times \mu_l^{-1} + \frac{\lambda_v}{\lambda_c} \times \mu_v^{-1}$.

The blocking probability, for an $M/G/s$ system, denoted by $B(s,a)$, where $a = \lambda_c/\mu_c$ is then given by ([5]) $B(s,a) = \frac{a^s/s!}{\sum_{k=0}^{s} a^k/k!}$.

Since all the three job streams are assumed to be Poisson, the blocking probability, B, for each of the three jobs is same and equal to $B(s,a)$. We also model re-attempts on blocking, by assuming that everytime a request is blocked, it is re-attempted with probabillity r. With re-attempts, the *effective* arrival rate, λ_{eff} of a stream which has arrival rate of new requests = λ, is given by $\lambda_{eff} = \lambda/(1 - rB)$. This gives rise to a set of interdependent equations, which can be solved iteratively until convergence is achieved.

The blocking probability for a "message" is then given by $B_{message} = 1 - (1 - B)^n$, where n is the number of fragments that a message is split into.

IV-C. Model Parameterization

In the absence of actual data from cellular operators, we "derived" various model parameters judiciously, as described below.

Assuming that about one million SMS messages [7] are sent and received in Mumbai daily by an operator and that there are about 50 cells, the number of SMS messages per cell per second is 0.55 (=$1000000/(3600 \times 10 \times 50)$) (assuming that most of the SMS messages are sent in 10 hrs during the day). Assuming that the number of voice calls are about 12 times[8] as many as SMS messages we have $\lambda_v = 6.6$. We simply assume that location updates are about 10% of this volume of SMS messages, giving us, $\lambda_l = 0.05$ per cell per second.

Since the GSM connect time requirement for voice calls is less than 4 seconds [4], we assume that the voice call setup part takes 0.5 seconds, i.e., $\mu_v^{-1} = 0.5$ on the SDCCH. In case of location updates, we assume an additional 0.1 second of use of the SDCCH (Section III), bringing the total to $\mu_l^{-1} = 0.6$.

Recall from the previous section that a control multiframe has 4 SDCCH slots for each user. The total data that can be carried on these four slots is 456 bits

(4×114 [4]). But for a signaling channel, 184 bits of data is encoded into 456 bits [2]. So, the 4 SDCCH slots carry 184 bits of actual data. An SMS message of size c characters, contains $c \times 7$ bits (7-bit ASCII encoding), to transfer which we need about $(c \times 7)/184$ control multiframes. Recalling that a control multiframe is 235.5 ms in length, the total time required for transfering $7c/184$ multiframes is approximately $9c$. So, the time needed to send an SMS message of c characters is $9c$ ms plus the setup time. We assume the latter to be the same as that for voice call setup (0.5 seconds).

V. RESULTS

Figure 4 (top) shows the blocking percentage of the message *vs* arrival rate of messages (starting at 0.55 messages per second), for different values of the number of SDCCH subchannels while keeping all other parameters constant (average SMS message size is assumed to be 70 characters, and retry probability is set to 0.3). The graph provides some useful insights. For example, when $s = 12$, the blocking increases by an order of magnitude, if the SMS volume quadruples (SMS traffic on peak days in India was reported to be four times average SMS traffic). We also observe that blocking reduces by four orders of magnitude when the number of channels increases four-fold.

Figure 4 (bottom) shows the blocking probability of the message *vs* average size of the message. The arrival rate was fixed at 0.55 messages per second. If average message size is c characters, this translates to $(c/160) \times 0.5$ as the SMS arrival rate. We can see the jump in blocking after 160 characters - this is clear from the definition of blocking of messages longer than 160 characters.

VI. CONCLUSIONS

We conclude from our work that the effect of increasing volumes and sizes of SMS messages on cellular call performance can be significant. SMS uses a resource that is shared by other very important control messages (especially, voice call set-up), and our "what-if" analysis suggested that control channels might indeed become bottlenecks if SMS volume continues to grow dramatically. The main work remaining is validating model parameters using data from cellular operators. An end-to-end delay analysis of SMS would
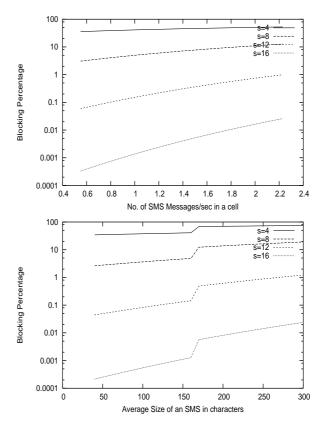

Fig. 4. Blocking vs SMS Arrival Rate and Average Message Size

also be an interesting study.

REFERENCES

[1] Asha Mehrotra, *GSM System Engineering*, Artech House, Inc., 1997.

[2] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks*, Pearson Education, 2002.

[3] *GSM 04.11: Point-to-Point (PP) Short Message Service (SMS) support on mobile radio interface.*

[4] V. K. Garg and J. E. Wilkes, *Principles and Applications of GSM*, Pearson Education, Inc., 1999.

[5] Robert B. Cooper, *Introduction to Queueing Theory*, CEEPress Books, Washington D.C., third edition, 1990.

[6] S. M. Redl, M. K. Weber, and M. W. Oliphant, *GSM and Personal Communications Handbook*, Artech House, 1998.

[7] Yahoo! India Finance, *SMS use rockets 300% on Diwali*, http://in.biz.yahoo.com/021105/26/1xdc2.html, 2002.

[8] Indus Mobile, *Why mobile markerting and advertising?*, http://www.indusmobile.com, 2001.