# Traffic Analysis of a Short Message Service Network


Vikrant Tomar, Himanshu Asnani, Abhay Karandikar, and Prateek Kapadia
TTSL-IITB Center for Excellence in Telecom (TICET),
Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India
Email: vikrant@ee.iitb.ac.in, himasnani@gmail.com, karandi@ee.iitb.ac.in, prateek@it.iitb.ac.in



***Abstract*—In this paper, we study Short Message Service (SMS) traffic in a nation-wide service provider network serving nearly 100 million messages a day. Our modeling approach is to study patterns in the service provider's data, and correlate these with well-known distributions. The network is modeled as a store-and-forward hub-and-spokes network - each message is a message between two spokes, with a switch at the hub deciding which spoke to forward the incoming message. This abstraction allows us to apply queueing models. A key contribution of this paper is the establishment of correlations between queueing models and the emprically analyzed data.**



***Index Terms*—Short Message Service, Traffic Analysis, Queueing Models, Arrival Process, Delay Distribution**


## I. Introduction

In the recent few years Short Message Service (SMS) has emerged as a popular means of communication between mobile users. Now a days, SMS, as an application, is not only limited to friendly text-messaging, but has grown into areas such as health care, customer care, and even rapid information dissemination purposes. In such a scenario, it has become crucial to assess the reliability of SMS network. This becomes particularly important as the SMS is also being considered for several mission-critial applications such as alerts and notifications in natural disasters or emergency situations [1]. For example, the service provider needs to be assured that the SMS delivery mechanism is not overloaded during popular festive events or disasters. Such an overload incident may be indicative of a system not provisioned appropriately.

There have been some work on the capacity analysis of SMS networks, for both CDMA as well as GSM systems [2], [3]. In [3], the authors have considered the SDCCH channel in GSM to be a bottleneck resource (since this is a common resource contended by location updates as well as control messages used in voice calls), and used a queueing loss model to obtain the SMS blocking probability. In [4], the authors have studied the empirical properties of SMS data at the message level and thread level. They have quantified various properties such as message length and call thread duration distribution.

In this paper, we obtain distributions of certain properties of the SMS traffic such as its arrival process, departure process, and service times. We have analyzed SMS data from a leading nation-wide service provider, and used a best-fit with well-known distributions. The data we analyzed was the Call Data Records (CDRs) for two days. The data was provided to us after anonimyzing the users' privacy information. Some information available in the CDR is originating User ID (an anonymous identification for a user), originating MSC, destination User ID, time of sending and delivering the message, and status of delivery. Some of the data statistics are listed in Table I. Unlike [3], in which the authors have modeled the system at channel level, we model the system at the Short Message Service Center (SMSC). By modeling the SMSC network, we have been able to study its performance parameters, and give some plausible recommendations to the service providers for a improved functionality of the network. Since only SMS traffic goes to the SMSC, we do not have any competition with messages used in voice call setup or location update.

TABLE I
**Some Statistics of CDR data and Service Provider Network**

| | |
|---|---|
| MSCs | $\approx$ 100 |
| A2P SMSCs | 2 |
| P2P SMSCs | 3 |
| STPs | 4 |
| ITPs | 4 |
| Number of Cells | $\approx$ 8000 |
| Average SMS Volume | $\approx$ 88 million/day |
| Voice/SMS Traffic | $\approx$ 2 |
| Buffer Length per SMSC | 0.6 million |
| Channels per SMSC | 30 (E1 TDD Link) |
| Message Delivery Attempts (MDAs) per SMSC | 4000/sec |

Rest of the paper is structured as following. Section II briefly explains the representative cellular architecture. Section III elaborates on the SMSC centered hub-and-spokes abstraction. Section IV provides and discusses various empirical analyses of the SMS traffic. Section V presents some key observations, and recommendations to the service providers. Finally, Section VI concludes the paper.

## II. SMS Network Architecture

In this section, we briefly describe the representative cellular system architecture, followed by SMS call flow, and state the respective assumptions.

### *A. Cellular System Architecture*

A simplified architecture of the representative cellular system is depicted in Figure 1, [5]. In the figure, *MO* refers to the mobile (or application) from where the SMS is sent. Likewise, *MT* refers to the intended recipient mobile (or application). For abstraction purpose, we have not included Base Stations in the architecture. A Base Transceiver System (BTS) exchanges message between the Mobile Switching Center (MSC), and a mobile user in its cell. The MSC network routes all SMS traffic through the Short Message Service Center (SMSC) network. The MSC is linked to Home Location Register

(HLR), and Visitor Location Register (VLR). The HLR is the functional unit that manages mobile subscribers by maintaining all subscriber information (e.g. electronic serial number, directory number, international mobile station identification, user profiles, current location). The HLR may be co-located with an MSC. One HLR can serve multiple MSCs, and also may be distributed over multiple locations. The VLR is also linked to one or more MSCs, and is the functional unit that stores and updates data related to mobility. When MS enters a new service area covered by an MSC, the MSC informs the associated VLR about a mobile by querying the HLR after the mobile goes through the registration procedure. The MSC are inter-connected though a Signaling Transfer Point (STP) and Internet Protocol Transfer Point (ITP) network. The STPs are connected to ITPs in a mesh topology by SS7 links, and the total traffic is equally divided between the ITPs. They balance incoming traffic load onto outgoing links to the SMSC. In addition, the network is also consisted of Data Message Handler (DMH), which is used to collect the data for billing purpose.

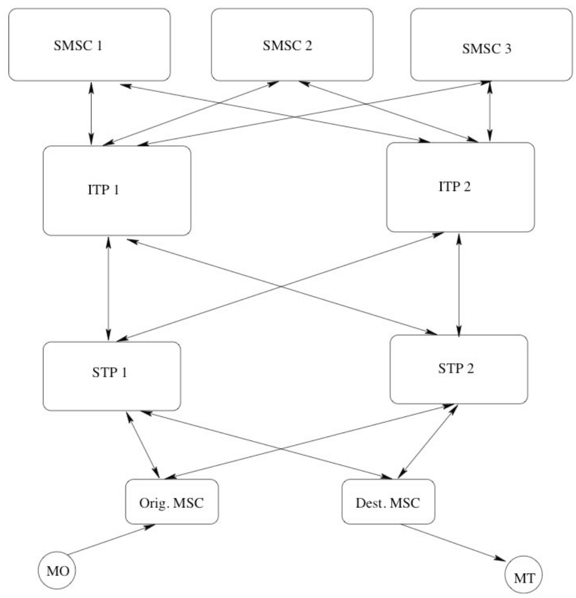


Fig. 1. Simplified architecture of the discussed SMS Cellular Network

### B. SMS Call Flow

The following steps detail the flow of an SMS through the service provider network.

1) When a user sends an SMS, it first goes to the base-station in the user's cell, and from there is forwarded to the Originating MSC. The MSC forwards messages as they arrive and may thus queue them for systematic forwarding. However, no processing is done at the MSC level.
2) The Originating MSC forwards the message to the STP-ITP mesh network, which essentially functions as an SMS switch.
3) The ITP then forwards the message to one of the three SMSCs in a round-robin fashion. The SMSCs function as store-and-forward buffers for the incoming messages and guarantee reliable delivery. It maintains a queue of the incoming traffic and serves messages in First In First Out (FIFO) fashion. The CDRs of the messages are also generated at the SMSC.
4) As an SMS arrives at the SMSC, it goes to the end of the queue and an Arrival CDR is generated. While serving a particular SMS, the SMSC queries the HLR to locate the destination MSC and then delivers the message directly to the MSC which then forwards the same to the mobile station through the respective base station. This entire cycle is defined as one Message Delivery Attempt (MDA), and an SMSC (in the analyzed service provider network) is capable of approximately 4000 MDAs/sec. If the HLR/VLR is not able to locate the mobile station then the SMS is queued in waiting. When the mobile comes online, this information is updated in HLR/VLR which then sends this information to SMSC. The SMSC then delivers the pending SMS. Pending SMS will be queued for a maximum duration of 24 hours, and delivery is retried for a fixed number of times.

## III. Single Server Abstraction of the SMSC Network

In our analysis, we model the SMSC network described above as a single message server. In this model, each message is tracked only from the originating MSC to the destination MSC. The SMSC processes each message before sending it to the destination MSC. This processing, in the real SMSC, corresponds to querying the HLR/VLR for the destination MSC point-code, delivering the SMS to the destination MSC, and book-keeping functions, like writing the CDRs. A message is queued by the server until messages received earlier complete processing. This abstraction is shown in Figure 2. This abstract model is only concerned with the SMS traffic, as voice call setup or location update messages do not go the SMSC.

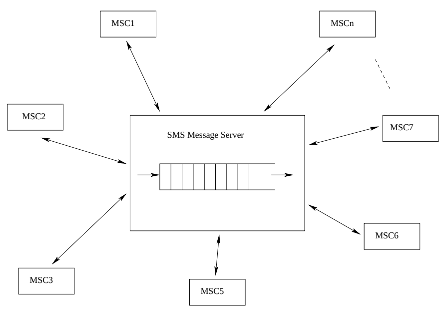


Fig. 2. Single server SMSC network abstraction

We are now ready to present the analysis of the SMS data.

## IV. SMS Traffic Analysis

### A. Message Length Distribution

In this network the maximum length of SMS could be 160 characters. If a message has more than 160 characters, it is sequentially chopped into messages of size less than or equal to 160 characters, and then each of these parts are delivered separately. A message is said to be successfully delivered only when all of the parts are delivered. Fig. 3 illustrates message length distribution. It can be observed, from the figure, that only few messages are large.

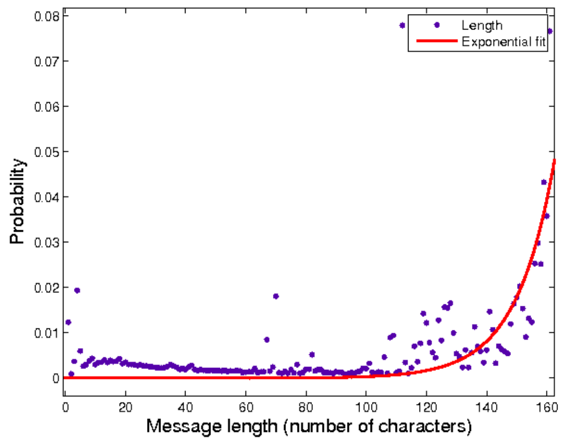


Fig. 3. Message Length Distribution

### B. Arrival Process

The network carries two types of SMS data, Application to Person or Person to Application (A2P) type, and Person to Person (P2P) type. In the network analyzed by us, different SMSCs are used for handling A2P and P2P type traffic. We have observed that roughly 60% of the traffic is contributed by A2P messages. Every SMS transaction is recorded as two different CDRs at the SMSC (a) Mobile Originating (MO) CDR, which stores certain information at the time of submission of the SMS to the SMSC, and (b) Mobile Terminating (MT) CDR, generated at the termination time, i.e., when SMS is delivered to the intended destination. We have parsed the MO CDRs for submit timestamps of SMSs to characterize the arrival traffic on the combined SMSC unit. Fig. 4 illustrates the cumulative arrivals as a function of time for one day period.

From Figure 4, we observe that, on an average, there is less traffic from midnight 0000 hours to 0800 hours in the morning compared to the late evening hours. Hence, to precisely characterize the traffic, we separately analyze data in windows of few hours, in which the plot in Fig. 4 exhibits linear behavior. As characterized by the CDRs, the SMS traffic has granularilty in seconds, hence to take into account the time resolution for the arrival process, we determine the distribution for increments in intervals ranging from 1 second to several seconds, and even minutes. From various simulations, we observe that such a distribution is a Gaussian process, as illustrated in Fig. 5 and Fig. 6, or sum of Gaussian processes, as illustrated in Fig. 7, with non-zero mean. This illustrate that arrival traffic is a type of Wiener process.

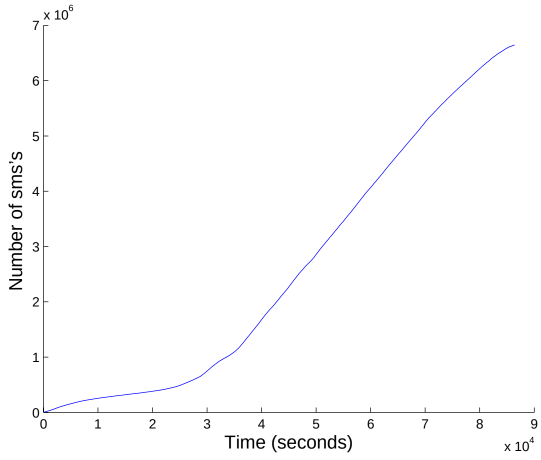


Fig. 4. Cumulative arrivals for data of Day 1

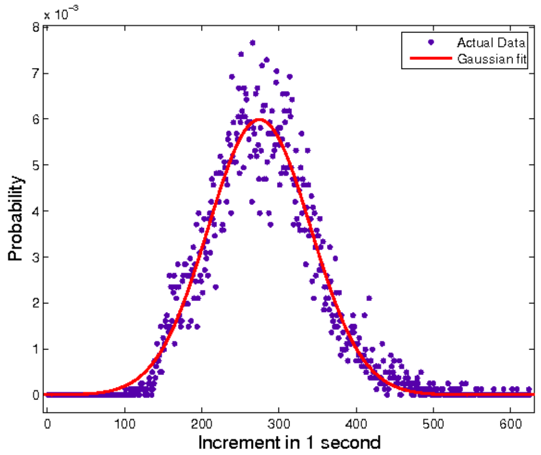


Fig. 5. Arrival Distribution for First 2 hours for Day 1

### C. Delay Distribution

An SMSC connects to the STP-ITP mesh network using SS7 E1 links, i.e., 30 channels each capable of rate $64\ kbps$. A single transmitted SMS can have upto 160 characters, where each character is encoded in ASCII format, and thus needs 7 bits. Therefore, an SMS of character length $c$ needs $7c$ bits (excluding overhead). Each SMS consumes one of the 30 channels of E1 link, hence the expected message transfer time is $7c/64000$, which is very small even for the maximum characted length, i.e., $c = 160$ characters. Therefore, in order to characterize the average delay, we can safely assume the message transfer times to be negligible. We have illustrated delay distribution, based on the time-stamps in SMS CDRs, in Fig. 8. Note that waiting times are exponentially distributed.

## V. Observations

The analysis in this paper leads to following observations for the service provider:

- There is less traffic from midnight 0000 hours to 0800 hours in the morning compared to the late evening hours.

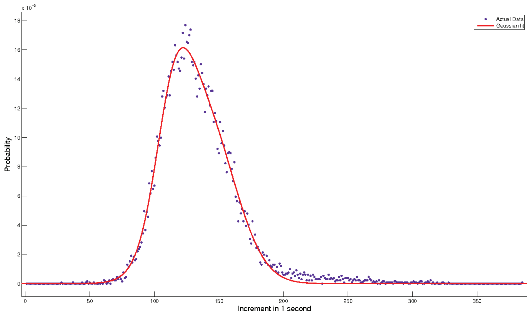

Fig. 6. Arrival distribution from 2 am to 4 am for Day 1

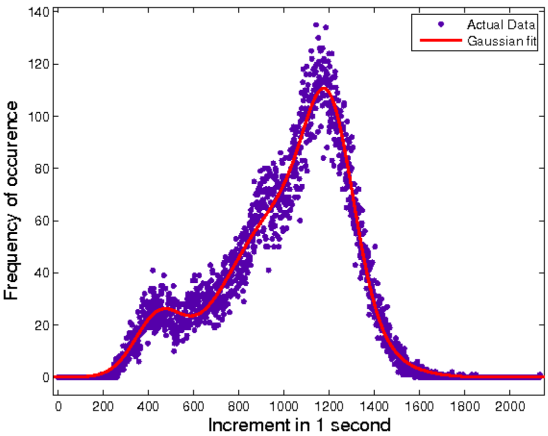


Fig. 7. Arrival distribution from 8 am to 12 midnight for Day 1

- From statistical studies on the SMS behaviors of the users, we saw that approximately 52% of the messages are exchanged within the same MSC. This information can be used for optimum network design. For example, we can employ an SMSC proxy at the MSC, to take care of the intra-MSC traffic.
- More than half of the traffic is contributed by A2P applications. Many A2P messages, however, have spurious destinations. This results in a waste of processing and computing resources as such messages are destined to be rejected. Statistics show that as high as 40% of the messages were rejected, which results in 24% of the total traffic. We recommend that the service provider should devise a scheme to minimize the messages sent to an invalid destination. An extra HLR lookup before sending the message may do the trick.
- Arrival process at both MSC and SMSC is Wiener, which has independent and identically distributed increments. However, unlike the Poisson arrival process in packet data networks, this is not memoryless. This indicates a strong correlation between past and present messages. This is expected since SMS "conversations" are common.
- Delay characterizes buffer occupancy. Exponential distribution indicates that queue is stable because probability of very high queue lengths is negligible. However, we advise the service provider to validate the traffic patterns at regular intervals to ensure scalability.

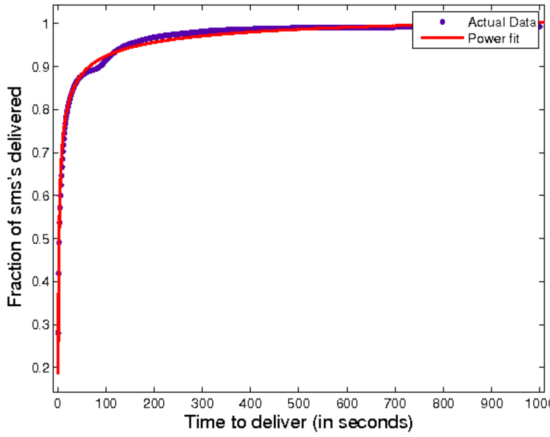


Fig. 8. Delay Distribution of the SMS traffic

## VI. Conclusion

In this paper, we have performed some empirical studies on two days of SMS traffic data of one of the leading telecom operators in India. The network is modeled as a store-and-forward hub-and-spokes network, where the hub refers to the SMSC, and the spokes to the links from MSC to SMSC (through STP-ITP mesh network). Each SMS travels through two spokes, with a switch at the hub deciding which spoke to forward the incoming message. We have analyzed the data based on this abstract model, and characterized traffic arrival process, message length distribution, and delay distribution. We conclude that, contrary to memoryless Poisson arrival in packet data networks, the arrival process in the SMS network is a Wiener Process, which indicates a strong correlation between the SMS "conversations". Furthermore, we suggest that an extra HLR lookup can be used to minimize the spurious A2P traffic, which contributes to 24% of the entire traffic. More than half of the total SMS traffic results from the messages exchanged within same MSC. The service provides can use SMSC proxy at MSC level reduce traffic load at the SMSC.

## References

[1] H. Rheingold, "SMS disaster warning system?" http://www.smartmobs.com/2004/12/31/sms-disaster-warning-system/, December 2004.

[2] A. Viterbi, "Capacity of a Simple Stable Protocol for Short Message Service over a CDMA Network," *Communications and Cryptography*, p. 423429, 1997.

[3] Leena Chandran, Nilesh Agarwal, and Varsha Apte, "Capacity Analysis of the GSM Short Message Service," in *Proceedings of NCC '03*, 2003.

[4] Petros Zerfos, Xiaoqiao Meng, Starsky H.Y Wong, Vidyut Samanta, and Songwu Lu, "A Study of the Short Message Service of a Nationwide Cellular Network," in *IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, New York, NY, USA, 2006.

[5] C. Smith and C. Gervelis, *Wireless Network Performance Handbook*, ser. Network Engineering. McGraw Hill, 2003.