Traffic Analysis for GSM Networks

M. Boulmalf[1], J. Abrache[1], T. Aouam[2], H. Harroud[1]
1:Al Akhawayn University in Ifrane, Morocco
2: Al Hosn University, AbuDhabi, UAE
e-mail m.boulmalf@aui.ma

Abstract— **When GSM was introduced in the early 90's, it was considered an over specified system. Nowadays, it is obvious that the whole range of services is widely in use. In addition, performance is degrading due to the rapidly increasing number of mobile subscribers, numbering over 2.9 billion subscribers around the world. The performance of cellular networks is the most important issue concerning their operators. The main goal is to keep subscribers satisfied with the delivered quality of service (QoS). In order to achieve the best performance, service providers have to monitor and optimize their network continuously. A Network Management System (NMS) with an online database is responsible for the collection of data on live networks. For greater effectiveness, operators install systems that do a lot more than collect and store raw data. These systems are easier to use and take advantage of all the data provided by the NMS. In this paper, we summarize measurements that were carried out on an operative GSM-1900 network to evaluate the performance of the GSM's Air-Interface (Um), during eight months. In this paper we have established statistically the following facts: (i) The peak hour in North America is between 4:00 and 5:00 PM. (ii) During week days the duration of calls increases from Monday through Friday. (iii) Weekend traffic is different from week-days traffic. Using a regression analysis we forecast traffic over time. The presented KPIs along with the derived statistical facts are crucial for operators who are concerned with maintaining a reliable and stable network, while maintaining an acceptable QoS.**

I. INTRODUCTION

GSM (Global System for mobile Communications) is the pan-European digital cellular standard published by ETSI (European Telecommunications Standards Institute). It is the most popular second-generation cellular system. In 2008, there were more than 2.9 billion GSM subscribers in more than 100 countries [1]. The GSM accounts for 71.5% of the World's mobile digital market. It demonstrates phenomenal customer growth running at twice the industry's predicted level. The GSM system architecture includes three standard interfaces: the Air-Interfac Um, the Abis Interface, and the A interface (Figure 1). The GSM functionalities are divided between the mobile stations (MS), base station subsystem (BSS), and mobile switching center (MSC) [2-4]. The BSS includes two types of elements: the base transceiver station (BTS), which handles the radio interfaces toward the MS, and the base station controller (BSC), which manages the radio resources and controls handovers. A BSC can manage several BTSs. Through the MSC, the GSM system communicates with networks such as the switched telephone network (PSTN), integrated services digital network (ISDN), circuit-switched public data network (CSPDN), and packet-switched public data network (PSPDN). In addition, GSM specifies three databases; the home location register (HLR), visitor location register (VLR), and authentication center (AUC). As a hybrid frequency-division time-division system, GSM organizes radio transmission by assigning carriers and time slots to logical channels. The frame duration is 4.615 ms, and each frame is divided into eight time slots. In section 2, performance of air interface was presented. Section 3 shows the experimental and the validation of results. Finally, section 4 presents the conclusion.

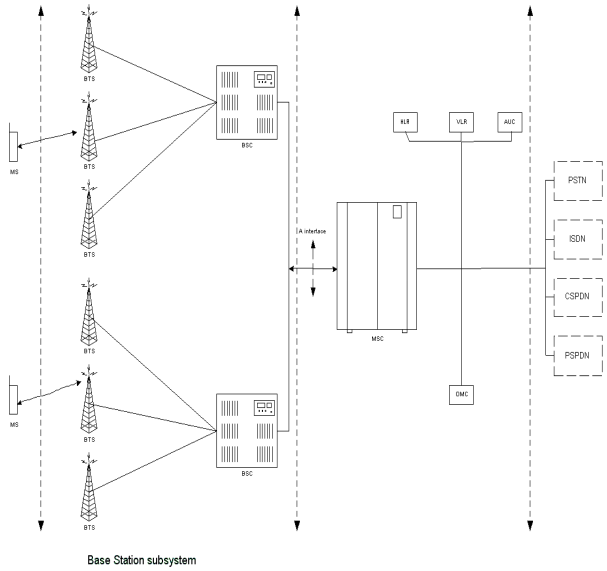

Figure 1. GSM system Architecture

II. GSM PERFORMANCE: AIR INTERFACE

There are several factors that can prevent a subscriber from making a call; for example, congestion or outage on any piece of the network (network unavailability). On the network side, three elements are related to congestion or indicate that a call could not be completed:

- Congestion on traffic channels, from a network point of view, this indicator is called TCH allocation failure;
- Congestion on control channels, from a network point of view, this indicator is called SDCCH allocation failure;
- Drop on control channels, from a network point of view, this indicator is called Call Establishment Failure.

A. Parameters and definitions

The following definitions are used for each cell at Peak Hour.

- Traffic: It represents minutes of calls in (Erlang). Traffic counters are registered in BSC's. There are two types of traffic that are used: Flow and Offered.
- TCH: A traffic channel represents a voice channel. Each call uses a TCH. There are eight channels defined (most for traffic channels and some for control channels) for each frequency.
- Percentage of TCH utilization: Number of minutes of calls that occurred at the same time over the maximum traffic permitted with a grade of service (GOS) equal 2%.
- Percentage of TCH congestion: It represents the percentage of the number of calls with congestion, no availability of traffic channel, divided by the total number of call attempt.
- Percentage of CCH congestion: It represents the percentage of the number of calls with congestion, or no availability of control channel, divided by the total number of control channel attempts.
- Percentage of CCH dropped: It represents the percentage of the number of dropped CCH divided by the total number of established CCH.
- Mean Holding Time: It is the mean time that a call last.

III. EXPERIMENTAL STUDY AND ANALYSIS OF RESULTS

In this section we present the results obtained through the analysis of operational data to help us comprehend the behavior and assess the performance of the traffic and logical channels. In GSM a new call cannot be initiated if SDCCH channels are not available and the same happens when SDCCHs are available but all TCHs are blocked [5,6].

Figure 2 depicts an overview of the system used for collecting raw data from the network elements. The raw data is extracted from the network elements, transformed and loaded into a data warehouse by making use of Business Intelligence (BI) tools. In addition to the extraction, transformation and load (ETL), BI tools provide historical, current, and predictive views of business operations, which bring greater business visibility and insight to the broadest audiences of users, allowing any user to have self-service access to up to the moment, relevant, and actionable intelligence. BI elements support reporting, interactive pivot-table analyses, visualization, and statistical data mining.

The data used in this experiment has been extracted from the test-bed network (Figure 3) which is composed of 1 MSC, 4 BSCs and many BTSs, over a period of 8 months.

Figure 4 shows the allure of daily voice traffic for weekdays and weekends. From the figure it is clear that the Peak Hour is 4:00 pm. In this curve we have two peaks; the first one corresponds to the lunch break, where there are a lot of calls. The second peak hour corresponds to the time immediately preceding the end working days. In this period many users start making calls. Table I shows the results obtained from the test-bed networks. The results are the peak hour of the network, the mean holding time, utilization, congestion, and drop calls. The selected performance indicators are very useful for service providers, since they allow them to take action very quickly and to add resources (radios) or perform cell optimization when needed. The test-bed shows that the average of the mean holding time is about 47 seconds. In this network each subscriber performs two handovers per call duration. Therefore the mean holding time from the network point of view is 94 seconds. The RF engineers can use this information to calculate the number of calls or Erlangs using this formula:

$$ErlangB = \frac{Established\_calls \times MHT}{3600} \quad (1)$$

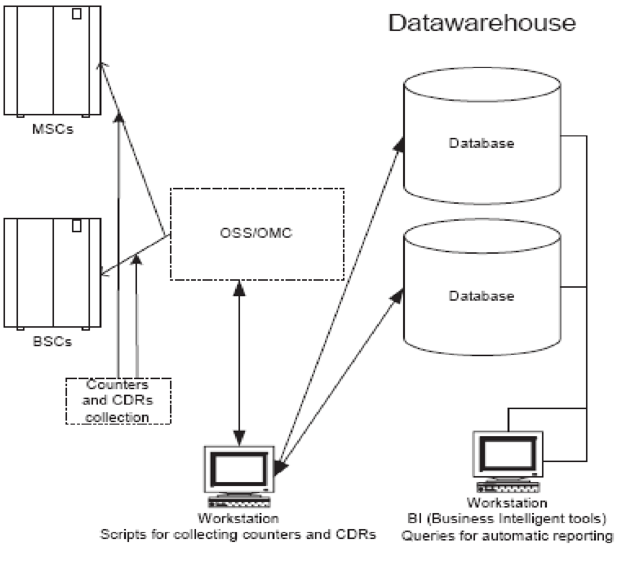

Figure 2. System for collecting brut counters and CDR's

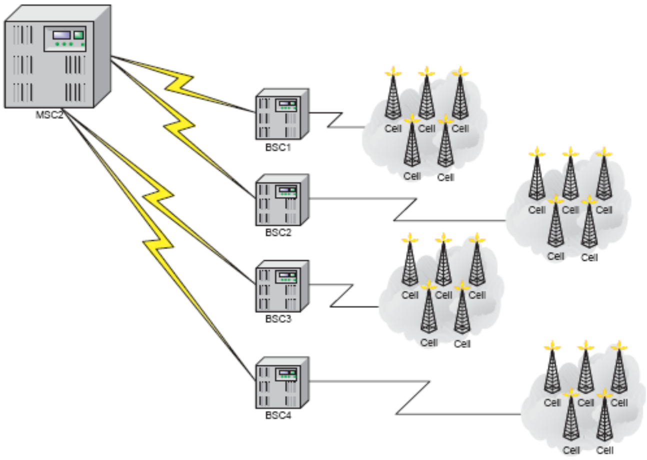

Figure 3. Testbed setup

TABLE I. KPI MEASUREMENT [8]

	MSC			
Network Performance	**BSC1**	**BSC2**	**BSC3**	**BSC4**
Peak Hour of the network	4:00 PM	4:00PM	4:00PM	4:00PM
Traffic during Peak Hour	7.70%	7.80%	7.40%	7.70%
TCH Mean Holding Time	46	43	57	46
Utilization (TCH)	61%	44%	45%	59%
Congestion on TCH	0.67%	0.10%	0.00%	0.62%
Drop call on TCH	0.54%	0.52%	0.86%	0.55%
Congestion on CCH	0.03%	0.00%	0.00%	0.03%
Drop calls on CCH	1.07%	0.46%	0.66%	1.02%

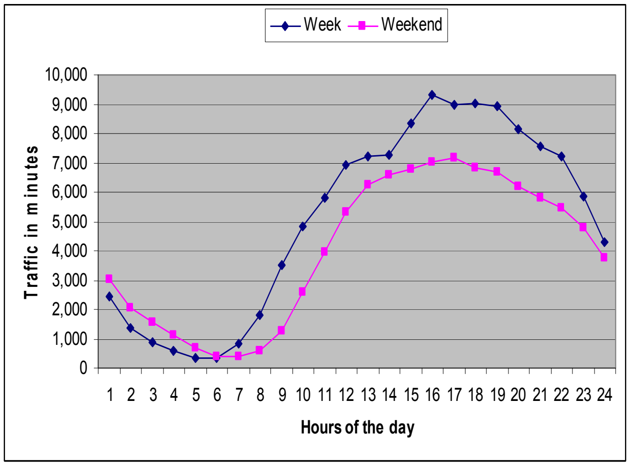

Figure 4. Traffic in minutes during week weekend days

It is noticed in Table I that the utilization is between 44% and 61%. This utilization indicates that the network is very well planned. In general the planning engineers have to add resources if the utilization reaches 80%. The congestion on traffic channels (TCH blocking Rate) are lower than the 2% threshold that is adopted by many operators in the world. The congestion on CCH must be very low, less than 0.1%. This congestion is very critical because if calls are blocked in this stage the operators lose big revenues. Utilization is the key indicator that characterizes a congestive scenario, as it is more representative of what is really happening and more independent from the situation itself than blocking. On the other hand, blocking provides additional information and thus, extra help in the identification and characterization of a congestion scenario. Blocking rate depends on certain network parameters, such as radio coverage of the selected area, overlapping of the cells and the use of directed retry [7].

The remainder of the numerical study examines three hypotheses, related to the traffic peak hour, the evolution of traffic volumes throughout the week, and the difference between the weekday and weekend traffic. In addition, it introduces a forecasting model based on regression that can be used to help planning engineers better predict traffic and add adequate resources to the network nodes when needed.

A. The peak hour in North-America is between 4PM and 5PM.

In this section we will prove statistically the timing of the peak hours. Figure 4 represents the measured average durations of calls for every single hour starting from the period 0:00AM – 1:00AM to the period 11:00PM-0:00AM. For convenience, these hourly periods will be referred to as hour 0 to hour 23 in the remainder. The hourly durations of calls are computed over the 243 days that constitute the time horizon of the study.

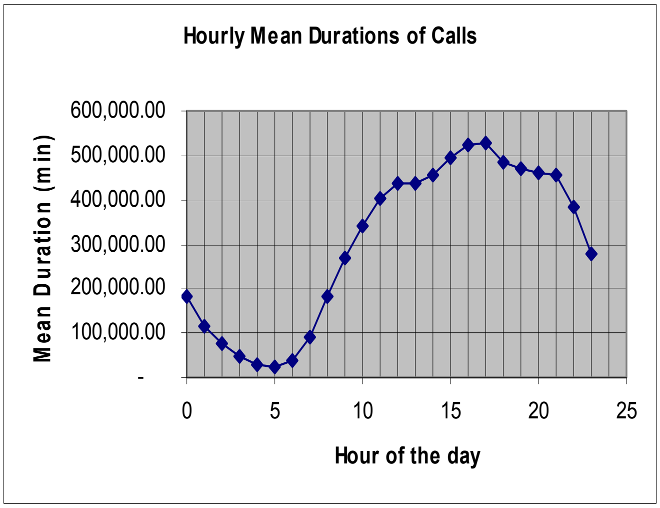

Figure 5. Cumulative Hourly Average Durations of Calls

Graphical inspection of Figure 5 gives the following general pattern for the durations of calls. The duration of calls pick up from their low point achieved at hour 5 (between 5AM and 6AM) and increase sharply to reach hour 12 (noon). The increasing trend continues, though at a lower rate, through the afternoon until hours 16 and 17 (between 4PM and 6PM), which constitute the daily peak period. The durations of calls then start decreasing, first slowly until hour 21 (between 9PM and 10PM), then at a sharper rate over the night.

In order to give more weight to this descriptive analysis in terms of statistical significance and gain additional insights into the hourly regimes, we conduct hypothesis tests for pair-wise comparison of successive hourly duration of calls. Each test will be a paired one (i.e., involving pairs of daily observations for each one of the hours being compared). More precisely, if μ_i indicates the mean durations of calls in hour *i*, then the test conducted to compare hour *i* to hour *i*+1 (hour 24 being equivalent to hour 0) can be formulated as follows:

$$H_0 : D_i = 0 \qquad (2)$$

$$H_1 : D_i \neq 0 \qquad (3)$$

Where $D_i = \mu_i - \mu_{i+1}, i = 0, \ldots 23$.

The results of the tests (two-tail p-values and the decision to reject equality of means or not at a significance level $\alpha = 5\%$ are given in Table II below. Rows in grey point out to a block of hours (12 and 13) that cannot be statistically distinguished from each other, in the sense that constitute together a local peak of traffic (such as hours, or hours 16 and 17).

TABLE II. COMPARAISON (T TEST) OF TRAFFIC IN SUCCESSIVE HOURS OF THE DAY

Test	Hours compared	p-value	Reject at $\alpha = 5\%$
H1	0 vs 1	1.407E-168	Yes
H2	1 vs 2	3.130E-169	Yes
H3	2 vs 3	8.323E-154	Yes
H4	3 vs 4	1.205E-89	Yes
H5	4 vs 5	5.718E-31	Yes
H6	5 vs 6	2.599E-53	Yes
H7	6 vs 7	8.537E-86	Yes
H8	7 vs 8	1.122E-112	Yes
H9	8 vs 9	4.326E-169	Yes
H10	9 vs 10	5.892E-137	Yes
H11	10 vs 11	1.101E-147	Yes
H12	11 vs 12	1.458E-81	Yes
H13	12 vs 13	49.65%	No
H14	13 vs 14	2.363E-42	Yes
H15	14 vs 15	1.796E-57	Yes
H16	15 vs 16	1.467E-44	Yes
H17	16 vs 17	0.79%	Yes
H18	17 vs 18	6.074E-55	Yes
H19	18 vs 19	3.554E-26	Yes
H20	19 vs 20	1.561E-11	Yes
H21	20 vs 21	0.01%	Yes
H22	21 vs 22	2.452E-138	Yes
H23	22 vs 23	2.137E-156	Yes
H24	23 vs 0	6.518E-106	Yes

B. During week days the duration of calls increases from Monday through Friday.

We conduct four hypothesis tests to compare the population average durations of calls in successive days (1-day lag). More specifically, let μ_j be the mean duration of calls in day $j, j = 1, \ldots, 4$ ($j = 1$ being Monday). The tests conducted are:

$$H_0 : D_j \leq 0 \qquad (4)$$

$$H_1 : D_j > 0 \qquad (5)$$

where $D_j = \mu_{j+1} - \mu_j, j = 1, \ldots, 4$.

Table III summarizes the one-tail p-values for the corresponding paired population t-tests, and the decision to reject the null hypothesis (i.e., there is a statistically significant difference between the days) or not at a significance level $\alpha = 5\%$. The results in the table reveal the existence of a statistically significant difference between all pairs of successive days from Monday through Friday.

TABLE III. COMPARAISON TEST) OF TRAFFIC IN SUCCESSIVE WEEKDAYS

Test	Days compared	p-value	Reject at $\alpha = 5\%$
D1	Monday vs Tuesday	0.03%	Yes
D2	Tuesday vs Wednesday	4.289E-06	Yes
D3	Wednesday vs Thursday	2.393E-09	Yes
D4	Thursday vs Friday	4.36%	Yes

C. Weekend traffic is different from week-days traffic.

The representation in Figure 6 of the average weekday (Monday through Friday) traffic and the average weekend (Saturday and Sunday) traffic clearly indicates a significant difference between the two types of traffic. To demonstrate that this difference is a statistically significant one, we conduct another paired t-test to compare the population mean of average traffic in weekdays with the population mean of average traffic in weekends. The result of the test is given in Table IV below and indeed proves the difference.

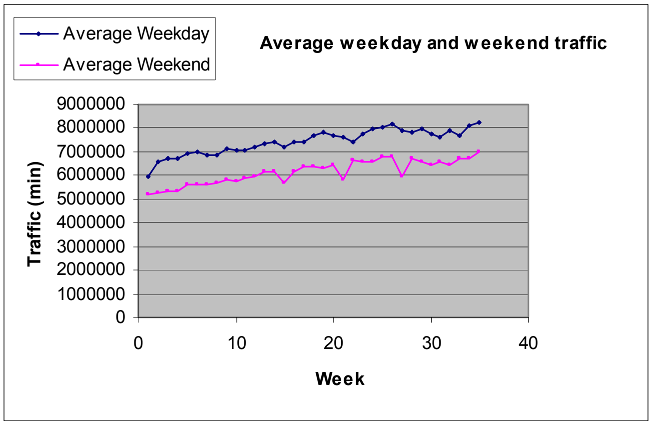

Figure 6. Average Weekday and Weekend Traffics Compared

TABLE IV. COMPARISON (TEST) OF THE DIFFERENCE BETWEEN WEEKDAYS AND WEEKENDS

Test	Comparison	p-value	Reject at $\alpha = 5\%$
W	Avg. weekday vs Avg. Weekend	0.03%	Yes

D. *Using a regression analysis we forecast traffic over time.*

Given the difference observed between weekday and weekend traffic, a simple linear regression model is inappropriate and has all chances to display a poor fit. We choose instead an exponential model frequently used for forecasting of seasonal time series. The model can be formulated as follows:

$$Y_i = \beta_0 \beta_1^{X_i} \beta_2^{Q_1} \beta_3^{Q_2} \beta_4^{Q_3} \beta_5^{Q_4} \beta_6^{Q_5} \beta_7^{Q_6} \varepsilon_i \quad (6)$$

Where

- Y_i: traffic in period (day), $i = 0, 1, 2, \ldots$;
- X_i: coded value of time; typically we simply take $X_i = i$;
- $Q_k, k = 1, \ldots, 6$: dummy variables (indicators) of the days of the week until Saturday; that is,
- $Q_1 = 1$ only for Monday, $Q_2 = 1$ only for Tuesday, etc., and $Q_k = 0, \forall k = 1, \ldots, 6$ for Sunday (taken as an arbitrary reference day);
- β_0: Y intercept;
- $(\beta_1 - 1) \times 100\%$: the daily compound growth rate of the traffic;
- $\beta_k, k = 2, \ldots, 7$: the multipliers of Monday, Tuesday, …, Saturday, relative to Sunday;
- ε_i: the irregular (random) component of the time-series.

This exponential model can obviously be linearized by applying a logarithmic transformation, which provides us with the following equivalent linear regression model:

$$\log Y_i = \log \beta_0 + X_i \log \beta_1 + Q_1 \log \beta_2 + Q_2 \log \beta_3 + Q_3 \log \beta_4 + Q_4 \log \beta_5 \quad (7)$$
$$+ Q_5 \log \beta_6 + Q_6 \log \beta_7 + \log \varepsilon_i$$

This model has been developed with the regression module in Excel. In terms of estimation error, the regression model has very good fit, with an adjusted R^2 of 84.4%, all predictors (independent variables) being globally and individually significant, and no serious violation of the linearity, independence, normality, and equality of variance conditions. The corresponding prediction equation to be used for forecasting can be written as:

$$\log \hat{Y}_i = 6.697 + (4.38\text{E-}3) X_i + (0.0947) Q_1 + (0.1218) Q_2 + \quad (8)$$
$$(0.1297) Q_3 + (0.1402) Q_4 + (0.1568) Q_5 + (0.0723) Q_6$$

IV. CONCLUSION

In this paper we have presented the results of an extensive study of a GSM network utilization. The experimental analysis focused on one particular metric – the duration of calls – and provided statistical evidence that: (1) the peak hour in a typical North American GSM network is between 4PM and 5PM; (2) the traffic increases through the week; and (3) there is a statistically significant difference between weekday and weekend traffic in terms of volume. The data collected has also been used to develop a regression-based forecasting model for traffic. These findings prove to be useful to network planning engineers as they provide them with decision making tools that help decide on timely and efficient investments on infrastructure. This experimental study can be easily extended to other GSM performance metrics and other wireless networks.

REFERENCES

[1] http://www.gsmworld.com/news/statistics/index.shtml
[2] M. Rahnema, ”Overview of the GSM System and Protocol,” IEEE Communications, April 1993.
[3] J. Cai and D.J. Goodman,”General Packet Radio Service in GSM” IEEE Communications Magazine, Oct., 1997.
[4] C. Bettstetter,H.J. Vogel, and J. Eberspacher,”GSM phase 2+ General Packet Radio Service GPRS: Architecture, Protocols, and Air interface” IEEE Communications Magazine, Oct., 1997.
[5] S. Kyriazakos, N. Papaoulakis, D. Nikitopoulos, E. Gkroustiotis,” Performance Evaluation of Operational GSM and GPRS Systems under Varying Traffic Conditions” IST Mobile and Wireless Telecommunications Summit, 17-19 JUNE 2002 Thessaloniki - Greece.
[6] S. Kyriazakos, N. Papaoulakis, D. Nikitopoulos, E. Gkroustiotis, C. Kechagias, C. Karambalis and G. Karetsos ”A Comprehensive Studyand Performance

Evaluation of Operational GSM and GPRS Systems under Varying Traffic Conditions ", IST Mobile Summit 2002, June, 2002, Thessaloniki, Greece

[7] T.S. P. Yum and K.L. Yeung, Blocking and Handoff Performance Analysis of Directed Retry in Cellular Mobile Systems, IEEE Transactionson Vehicular Technology, Vol. 44, No. 3, pp. 645-650,August1995.

[8] M. Boulmalf, A. Shakil , "Performance Evaluation of Operational GSM's Air-Interface (Um)", in Proc. of Applied Telecommunication Symposium," Orlando, Florida, pages 62-65, March 2003.