

MINING OPERATIONAL DATA FOR IMPROVING GSM NETWORK PERFORMANCE

Antonio Leong, Simon Fong

Department of Electrical and Electronic
Engineering
University of Macau , Macau

Edison Lai

Radio Planning Networks Division
Hutchison Telephone Limited
Macau

ABSTRACT

Operation and maintenance system (OMS) is an important part of network management in GSM networks. A large amount of operational data is generated daily in a mobile communication system, which contains hidden valuable information about the system behavior. The knowledge discovered from mining on such data can be used to analyze network performance and aid in long term planning. In this paper, we devise a solution procedure for mining correlated patterns from the operational data of a GSM system that come in a variety of formats. The results will be exploited for finding the critical factors that will affect the network quality and performance. Furthermore, it is believed that when a set of threshold values (that are used in OMS) are predicted in real-time by using a fuzzy classifier system from the rules generated from the past cases, the grade-of-service could be improved while maintaining a minimum number of cells.

1. INTRODUCTION

Telecommunications has a tremendous growth over the past decade. In this competitive market the robustness of the system, the network coverage, and the quality of service have become the key elements in retaining customers. It is important that the network performance is constantly monitored in order to ensure a high quality service being offered to the customers. In most mobile service industries, the benchmark of GSM network performance is mainly based on the following factors: Grade of Service (GOS), Drop Call Rate (DCR), Call Setup Failure Rate (CSFR) and the traffic statistics. Specifically, for monitoring a GSM system, a tool called Operation and Maintenance Control System (OMS) in OMC is used to probe the data of the above-mentioned performance criteria by using standard query data retrieval methods. The OMS compares the collected data against some threshold values, checks for any violation and proceeds to some recovery steps if needed. A drawback of

OMS is that the thresholds are fixed and predefined values which do not adapt to the actual situation. Likewise for GOS, it is calculated from a single cell instead of an aggregate value or a set of values representing the whole network. In essence, the standard queries that select data from cells site used by OMS nowadays are not enough to give an idea or any predicting indicator to the network engineers.

2. PROPOSED SOLUTION

In this paper, data mining techniques will be applied to explore the critical factors that will affect the network quality and performance. A set of threshold values from any outstanding cells will be analyzed in real time using a fuzzy classifier system. The objective is to optimize the current network performance by increasing the GOS offered to customers while maintaining a minimum number of cells. In particular, we will try to discover abnormal patterns that will directly affect the GOS. We start by analyzing current network performance and try to predict useful results by performing inference on the current data. The results collected can be used to optimize the current configuration. In section 3, we will briefly describe the technological background of the network performance measurement equipment. In section 4, we develop our solution for mining useful knowledge from a large set of raw data. Then we will elaborate on how we mine the operational data.

3. TECHNICAL BACKGROUND

The operations and maintenance system (OMS) provides the capability to manage a GSM network remotely. The OMS comprises of two parts: Network Management Center (NMC) & Operations and Maintenance Center (OMC). The NMC has a view on the entire PLMN (Public Land Mobile Network) and is responsible for the management of the network as a whole. The NMC resides at the top of the hierarchy and provides global network

management. The NMC offers the ability to provide hierarchical regionalized network management of a complete GSM system. It is responsible for operations and maintenance at the network level, supported by the OMCs which are responsible for regional network management. The NMC is therefore a single logical facility at the top of the network management hierarchy. The functionality of the NMC can be as follows:

- 1) Monitor nodes on the network.
- 2) Monitor GSM network element statistics.
- 3) Monitors OMC regions & provides information to OMC staff.
- 4) Passes on statistical information from one OMC region to another to improve problem solving strategies.
- 5) Enables long term planning for the entire network.

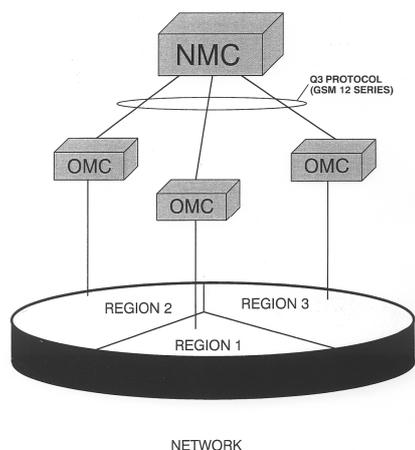


Figure 1 - Operations and Maintenance System

The OMC provides a central point from which to control and monitor the other network entities (i.e. base stations, switches, database, etc.) as well as monitor the quality of service being provided by the network. The role of the OMC within the GSM network management hierarchy is shown in Figure 2.

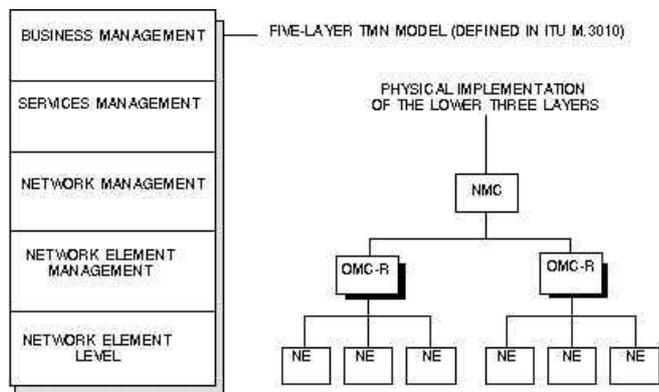


Figure 2 - Network Management Architecture

Figure 3 shows a functional representation of performance management data flow. Performance data is generated at the BSS and the RXCDR, and stored as a performance measurement (PM) data file. Performance measurements include traffic measurements within the telecommunications system (both user and signaling traffic), quality of service measurements (delays in call setup) and availability measurements (the beginning and end times of service unavailability). At 30 or 60 minute intervals (defined by the user), the PM files are sent to the OMC. The OMC processing software (known as the parser) then converts the data in the files into raw statistics suitable for storing in the INFORMIX database at the OMC.

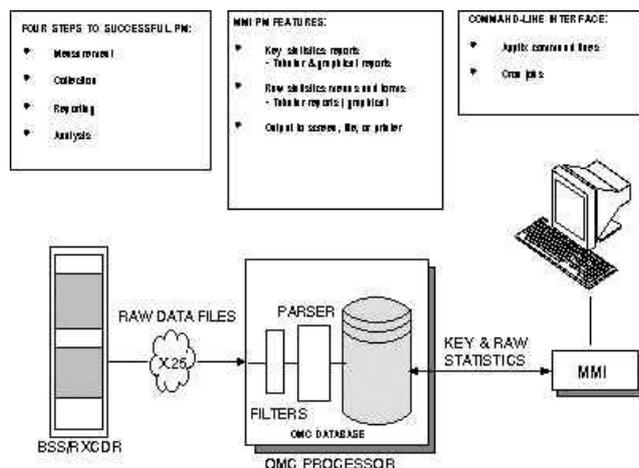


Figure 3 - Performance management

Measurement results generated in Network Elements (NEs) are sent to the OMC as raw data via an X.25 network. The collected data files are processed into raw statistics and stored in the OMC database. The data files collected are processed into raw statistics. These raw statistics are used for creation of key, network health or custom statistics. These statistics might include call processing, interface, and processor utilization measurements.

Key statistics are generated at the OMC by processing raw statistics generated by the Base Station System (BSS) using some predefined algorithms. These statistics are designed to give an overall indication of the condition of the system and allow comparisons of similar time periods over a span of time, to help detect congestion trends and possible performance degradation. Key statistics also provide means to facilitate the monitoring the most important network parameters. For example, various handover failure statistics may be combined and averaged over the total number of calls, to produce a handover failure rate key statistic.

Network health statistics are calculated at the OMC using a combination of raw and key statistics. These

statistics are used to create BSS network performance reports. These reports provide an indication of the networks health from the subscriber's perspective.

Traditionally, key and network health statistics are analyzed manually by a technical group to discover irregular patterns in order to measure current network performance. In this paper, we will try to analyze key statistics data using data mining techniques in order to obtain useful patterns for measuring network quality. Specifically, we wish to see some sequential and correlated information among all the parameters.

The results should be able to give some new insights of network performance, and can be further analyzed to provide operators information needed for network management functions, such as quality-of-service analysis or planning the evolution of the Public Land Mobile Network (PLMN). Next we will describe briefly the key statistics factors affecting the network performance.

4. DATA-MINING ENVIRONMENT

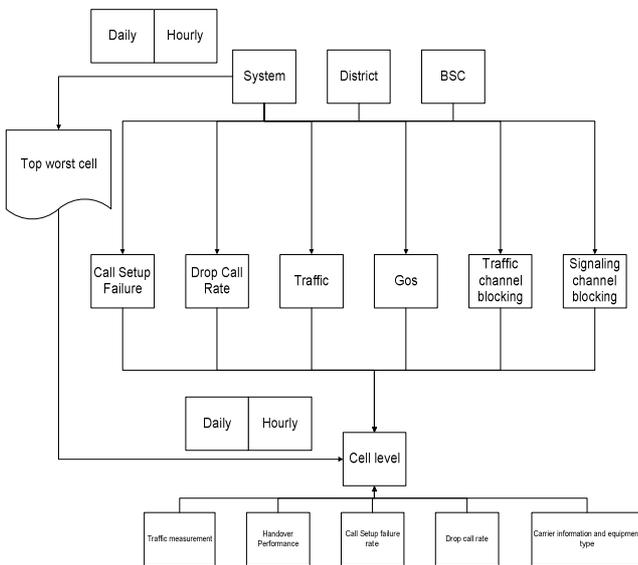


Figure 4 – Flow of data for data mining in OMC

The following figure shows the relevant key statistical data collected and processed by OMC. Data are collected in hourly and daily basis at regional level or cell level. Technicians will first analyze system-wide data in order to discover unusual patterns. They will have to discover the cause of irregular pattern by investigating the origin of unusual network behavior. Once the search region is reduced, cell level data are analyzed to discover unusual pattern affecting the overall network performance. We have been applying a data-miner called Castaneda and fine-tuning the appropriate data-mining algorithms for identifying patterns that have impact on GSM network performance.

Consequently we have defined several categories for the data mining results that are obtained from different data mining methods in order to discover interesting knowledge from our data. They include visualization, clustering, association rule and association rule for classification. We use map visualization to maintain daily base station performance to see if any abnormal event occurs. The following mining result based on historical data can search some useful and interesting pattern, trend, information etc. This knowledge can help in dimensioning the network loading and maintain the network performance.

4.1. Traffic Analysis by Visualization

Since the radio coverage of GSM networks are constructed by different base stations, it is needed to monitor the performance of each cell site daily. In this section, we have mapped parameters such as Traffic, Tch Gos, Csfr and Dcr by geographical means, i.e., it distribution along Macao. Also, we can assign other attributes for displaying the result on a geographical map. For traffic, it shows the distribution of traffic density in different areas. For the Tch Gos, it shows which area has higher Tch Gos.

These visualization results (shown in Figure 5) can be used for traffic dimensioning and resources management, and it can easily alert traffic problems and aid in shifting some high traffic area to other area (from Red to Green) in order to balance traffic loading. For the Csfr and Dcr, those areas with red or orange spots are areas with high call setup failure and drop call rate. By alerting this, network operators can design a better network and ensure good quality and performance for the base stations.

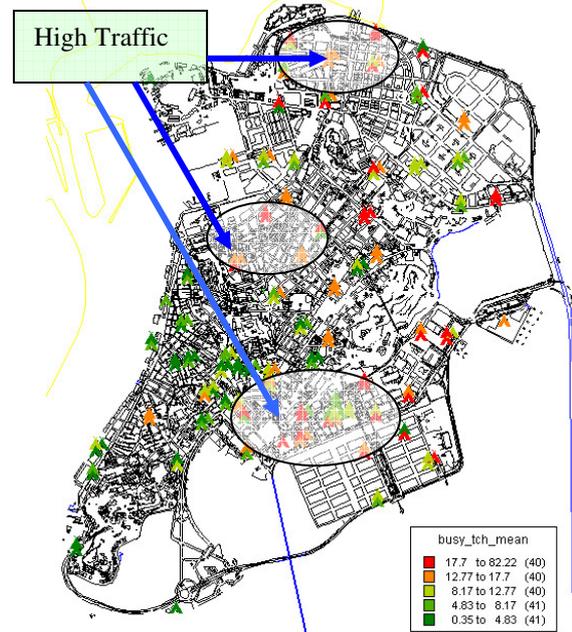


Figure 5 - Busy_Tch_Mean distribution on a visual map.

4.2. Traffic Analysis by Classification

We start by aggregating weekly data on weekday basis and try to find the distribution of several parameters over each day of the week. Castaneda classifies each parameter into five groups with simplified visualization process. In particular we try to analyse the traffic, Tch blocking, and Sdcch blocking over each day of week in order to gain a deeper understanding of high traffic and blocking dates. By gaining access to this type of information, we can better allocate our resources and equipments in keys dates such as days of week with high traffic.

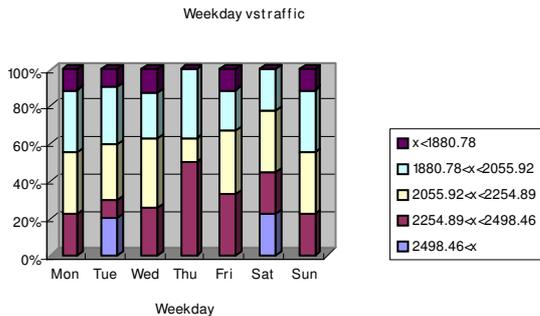


Figure 6 - Traffic distribution by weekday.

Similarly, we can aggregate traffic, TCH blocking and SDCCH blocking results in order to find regions with high traffic and blocking rate. For example we can show traffic density allocated in each district and its levels are as follow: Highest in Macau North and Macau Central, Medium in Macau South region and Lowest at Coloane and Taipa.

For checking the Tch blocking, the data is drilled down into district level, showing that the highest Tch blocking is allocated in Macau North region and Taipa. Hence it is needed to reconfigure the capacity of these two districts. In general, the usage of this information is for network planners to optimize current traffic capacity by district, by day/hour, by Qos, by equipment and by interference.

4.3. Traffic Analysis by Clustering

By using Castaneda, we have defined three clusters, with number of iterations equals to 50. Generally, the data are successfully clustered within 20 iterations. Using clustering, it gives us insight of the data when there are too many input attributes. This helps to clarify the healthy status of the system at different levels, such as: system daily, system hourly, district daily and district hourly level.

By analyzing the input data, we can have a general idea of the status of the system. For example, there is 5.88% of DCR (drop call rate) between the ranges of 5.14 to 5.73 in Cluster 0. By notifying this result, one can use this information to perform a simple query on the raw data to

find out cells within this range to detect the problematic cells causing high DCR. A sample of such cluster is given in the appendix.

4.3. Association Rule & Association Rule for classification

In our experience it is not wise to mine with association rule without a target attribute because it will give a lot of different combination of association rules. Using association rule is suitable when the data source does not contain more than 10 columns. In our project, we try to find useful rules that control the network performance. We try to guarantee the main attributes Call setup failure Rate (Csfr), Drop Call Rate (Dcr), Handover Drop Call Rate (Dcr Ho), Handover Failure Lose (Ho Fail Los), Handover Failure and Recover (Hof) and Traffic channel Grade of Service (Tch Gos) under a threshold value. Using Association Rule for classification, it can give some general rules on how to maintain the network performance.

Csfr Rule Classifier

In radio technology, CSFR happens due to lack of Tch (Tch blocking) and Sdcch (Sdcch Blocking). If the Csfr is due to lack of Tch, expanding the current number of channel by adding extra carrier can relieve the problem. However, if Csfr is a result of lack of Sdcch, it is needed to decrease the signaling loading of a Base Station (BTS). For example, both Hand In (Ho_per_call) and Hand Out (o/i_inter_bs_ho) will use resources of the signaling path. It is difficult to add extra carrier for Tch and Sdcch due to limited resource and frequency planning. Another reason that causes CSFR is due to Interference (Ioi, crmsfail), making the signal and information impossible to be transmitted correctly through BTS and Mobile Station. Beside of this, Csfr may be related to different BSC (limited resource or fault of Switching Center). Finally, there is a grading named as GOS (Tch Gos) for each cell, that is related to its current number of TCH and Traffic.

Dcr Rule Classifier

High Dcr (Drop Call Rate) is usually due to the interference (Ioi, crmsfail). Since Dcr is related to the rating of Dcr Ho, we have to decrease Dcr_Ho (Drop Call During Handover) and Ho_fail_los (Handover Failure and Lost) to maintain Dcr at a certain level. If there is any Sdcch blocking, this will cause out of resources problems and causing problems in process the signaling information proceeding. A lot of handover will also cause BTS overloading in handling the handover process (i/o_inter_bs_ho, ho_per_call). In order to maintain both Dcr and Dcr Ho under the prescribed threshold value, it is needed to maintain certain performance value under classified level.

Ho fail los & Ho Fail and Ret Rule Classifier

The most useful rules for Ho fail los & Ho Fail and Ret are its relation to Ioi_mean, Sdcch blocking, Ho_per_call. The relation with Tch_gos and Tch_blocking just show the performance of traffic distribution. First of all, to keep IOI in healthy status, we have to ensure the frequency is not in interference with others by keeping Sdcch blocking in a low level. This can ensure that a cell has enough resource to handle the handover command. For Ho_per_call, if we can keep this value at lower levels, it can be assumed that unstable handover problem is cleared. It's not usual for a phone call to handover over than three or more times when it is serving in a stable and good coverage area.

5. CONCLUSION

Traditionally the network statistics are monitored by a system-calculated figure and the irregularities are analyzed manually by a group of technicians. By mining the

operational data the hidden patterns and irregularities are better revealed. Datamining is not only served as a tool for network experts to analyze the network performance but also provides hints for future network failure incidents given the past result generated by data mining.

By means of map visualization the traffic channel monitoring is more obvious and remarkably noticeable at a glance. Traffic analysis by classification provides the network engineers the usage patterns over each slot of timeband, dayband versus others attributes such as district, cell handover etc. Clustering allows the data grouped by its similarity and speaks of itself. Preventing from the perception or "do it as usual practice", network engineers could observe the change of patterns of the data through clustering. Association provides the inter-relationship of the factors and gives hints of controlling. For example, we can keep the DCR at certain acceptable rate by controlling the traffic channel at certain levels.

Attribute	Values	Cluster 0	Cluster 1	Cluster 2
weekday	Friday	0.176471	0.115395	0.157895
	Monday	0.058824	0.192308	0.157895
	Saturday	0.352941	0.038462	0.105263
	Sunday	0.352941	0.038462	0.105263
	Thursday	0	0.192308	0.157895
	Tuesday	0.058824	0.289231	0.105263
	Wednesday	0	0.153846	0.210526
Csfr_discr	4.29910628019324_x_4.58421739130435	0.411765	0.346154	0.368421
	4.58421739130435_x_4.87733333333333	0.235294	0.038462	0
	4.87733333333333_x_5.14583333333333	0.058824	0	0.105263
	5.14583333333333_x_5.7325	0	0.076923	0
	5.7325_x	0.058824	0.038462	0
	x_4.29910628019324	0.235294	0.5	0.526316
	Dcr_discr	1.36116883116883_x_1.43906565656566	0.294118	0.423077
1.43906565656566_x_1.51861111111111		0.529412	0.038462	0.421053
1.51861111111111_x_1.61		0.117647	0	0.210526
1.61_x		0.058824	0	0.052632
x_1.36116883116883		0	0.538462	0

Figure 7 – An example of system cluster.

Description			
This statistics is the proportion of MSs which, having successfully accessed the TCH, subsequently experience an abnormal release, caused by either radio or equipment problems. This includes RF losses and losses during handovers.			
Usage: Fault finding, Service retainability; Basis: Cell, BSS or network			
IF	ioi_mean1_discr=x_-108.475133161625	THEN	dcr_discr=x_0.599939752708616
IF	crmsfail1_discr=x_169.017459661442	THEN	dcr_discr=x_0.599939752708616
IF	dcr_ho_discr=x_0.39818242712727	THEN	dcr_discr=x_0.599939752708616
IF	ho_fail_los_discr=x_0.690756289456244	THEN	dcr_discr=x_0.599939752708616
IF	tch_blocking_discr=x_1.04035033488944	THEN	dcr_discr=x_0.599939752708616
IF	tch_gos_discr=x_0.420142192932043	THEN	dcr_discr=x_0.599939752708616
IF	sdccch_blocking_discr=x_0.979319235616507	THEN	dcr_discr=x_0.599939752708616
IF	i_inter_bs_ho_suc_discr=x_48.3739023590557	THEN	dcr_discr=x_0.599939752708616
IF	o_inter_bs_ho_suc_discr=x_49.1075265041182	THEN	dcr_discr=x_0.599939752708616
IF	bsc_no=4	THEN	dcr_discr=x_0.599939752708616
IF	ho_per_call_discr=0.623648461325201_x_3.09738034538317	THEN	dcr_discr=x_0.599939752708616
IF THEN	dcr_discr=0.599939752708616_x_1.49461912977244		

Figure 8 – An example of Dcr Rule Classifier.