Data Mining for Resource Planning and QoS Supports in GSM Networks

Simon Fong
Department of Computer and Information Science
University of Macau
Macau SAR
Email: ccfong@umac.mo

***Abstract*—Applications that run on mobile phones shaped a trendy lifestyle for many users nowadays. This led to a significant growth in the proportion of data traffic, relative to voice traffic, to be delivered in the mobile phone network such as GSM. Traditionally the underlying radio resources in GSM networks for data and voice traffic were allocated by some predefined traffic policy which was manually configured. The allocation may not be most accurate for the fact that demands for data traffic fluctuate largely and temporally. A new resource planning scheme is desired that can dynamically adjusts the resource allocations according to the latest information of the traffic statues. In order to facilitate such dynamic resource allocation, a resource management system is proposed in this paper. Data mining is used to derive rules and extract traffic patterns that reveal critical information for setting values in resource planning. Empirical testing data are used in experiments that demonstrate the efficacy of the data mining techniques.**

***Index Terms*—Data mining, GSM network, Radio resource planning.**

I. INTRODUCTION

Many value-added services such as Web surfing, multimedia functions, mobile commerce and a wide variety of users' applications (e.g. Apple's iPhone Apps) ride on the data transmission services in GSM networks. In addition to providing quality service for voice services, telecommunication operators increasingly invest large sums into data transmission service provision that has a huge consumers market and it is still in the bloom. From the technical perspective of network management, it is therefore essential to minimize the occurrence of packet losses and delay, and to improve the quality of packet transmission services (QoS) in the network. For instance, mobile commerce [1] that leverages on data or message transmission has a stringent set of performance requirements on the delivery of messages, such as loss and latency. Network quality is a core business concern.

Given a limited capacity of resources offered by the fixed hardware infrastructure, a mobile network is statistically allowed to relocate its radio resources in proportions based on the forecast of the users' demands that usually is specified in a traffic profile. Some issue of radio resource management and congestion that needs to be solved includes the concentration of large number of users or requests in a hot spot area causing an upsurge.

Radio resource mobility prediction has hence become an important topic in mobile network management [2]. In this paper, we show that how data mining algorithms are being applied for predicting and managing the radio resource movement of cell. Cell usage and performance are mined from the history of operational data, and then the prediction rules are extracted from these patterns. Radio resource management and reconfiguration then follows by using these rules. In this way, the network resources can be dynamically adjusted to meet the ever-changing demands and most-updated usage from the population of mobile users on a daily basis.

In the literature, a lot of research works have been done on radio resource management in mobile networks in the context of predicting the traffic loads across different cells by tracking the movements of mobile phone users. The radio resource prediction is about predicting a cell next usage movement, where the cell is serving for different mobile users traveling in and out [3]. The predicated movement can then be used to increase the efficiency and utilization of mobile network by possibly shifting the resources in the cell neighborhood of a mobile user in advance for effective resource utilization and reduction in the latency of accessing the resources.

The resource planning approach and QoS support that we introduce in this paper, however takes a completely different scale. Instead of tracking the users' movements in each individual cell which obviously would be very computational costly, we periodically collect and analyze the operational data from the whole network at the Network Management Center (NMC), for identifying the so-called troubled spots in various regions of the network. Patterns and rules were recognized that led to those undesirable conditions by data-mining, hence radio resources would be preemptively allocated according to the optimized parameters in the traffic policy profile. Our approach should be relatively more practical because GSM network usually has a NMC infrastructure in-built and radio planning is a standard feature. See Figure 1.

This paper is organized as follow. Section 2 introduces the shortcoming of the current radio planning approach, mainly grounded from the manual setting of parameters that leads to under- or over- utilization of radio resources. In section 3 we propose a data-mining solution for resolving this problem. Section 4 showcases some results of mining GSM network data, followed by a conclusion.

doi:10.4304/jetwi.3.2.100-109

II. Performance Monitoring in GSM Networks

A. The Monitoring Infrastructure

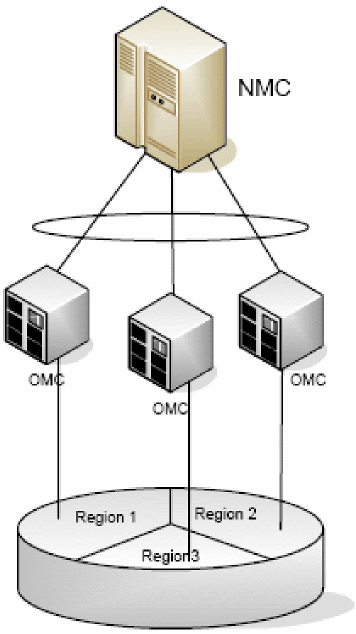


Figure 1. NMC and Operations and Maintenance System

In the heart of a GSM network, there is an Operations and Maintenance System. The system enables network engineer to configure and maintain the GSM network from a central location. It has mainly two parts: Network Management Center (NMC) and Operations and Maintenance Center (OMC). 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 of the whole GSM network system. As a single logical facility at the top of the network management hierarchy, it monitors real-time network system performance by checking for abnormalities, from base station faults, alarms to power supply problems. Instead of monitoring a particular cell, the whole system is examined, which might be composed by over hundreds to ten thousands of cells.

Each OMC under the NMC collects and stores performance statistics over the corresponding network element in the network. These statistics might include call processing, interface, and processor utilization measurements. They 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. These statistics are then used in the NMC for network performance analysis and aid in long term planning. More details about NMC, OMC and the statistics can be found in [4].

Network health statistics are calculated at the OMC using a combination of raw and key statistics. Currently, the measurements of the GSM network performance are mainly based on several factors: Gos (Grade Of Service), Dcr (Drop Call Rate), Csfr (Call Setup Failure Rate) and traffic. The statistical results are formatted into reports that provide an indication of the networks health from the subscriber's perspective. Traditionally, such key and network health statistics are analyzed manually by a group of technical personnel in order to discover irregular patterns in order to measure current network performance.

B. Radio Resource Optimization

In general GSM network design, radio frequency planner often uses a constant Gos value such as 2% or 5% in planning the capacity of radio network. The arbitrarily chosen Gos value may not be best suitable for planning a radio network capacity that has different demands and requirements over time. Instead of a predefined value we opt to find a most fitting value for Gos by data mining. With a best fitted Gos, other traffic channel parameters such as Csfr, Dcr, Handover drop rate (Hod) etc, can be estimated for maximizing the radio capacity and utilization while keeping the network in a healthy condition. The objective is to optimize the current network performance by increasing the Grade of Service (Gos) offered to customers while maintaining a minimum number of cells. This can be achieved by finding a compromise between the resources for voice and data traffic, which is balance between the number of traffic channel (Tch) and signaling channel (Sdcch). Practically this could become possible by installing a resource controller server, called Radio Resource Management Server (RRMS) in the existing GSM network along with the NMC. The function of the RRMS is to optimize the system traffic loading for different radio channel configurations in GSM cell sites. The fundamental question that arises in the decision of system implementation would be: How many channels or resources are available for each site/system by the RRMS?

III. Our Proposed Solution

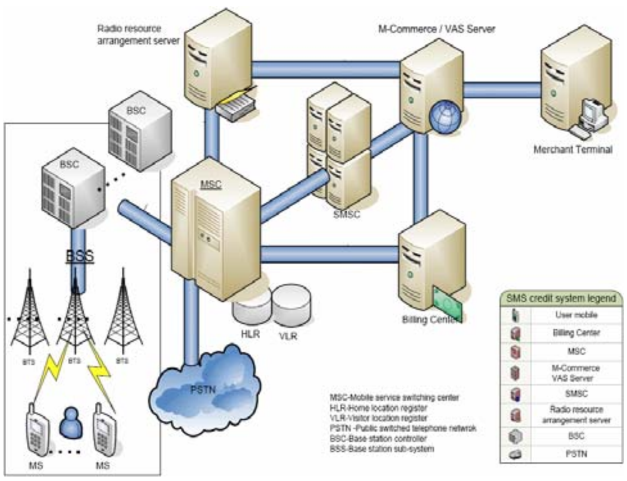


Figure 2. GSM Network Model with RRMS

A. *RRMS and Traffic Profiles*

In order to optimize and plan the appropriate radio resources for both voice and data service in GSM network, RRMS facilities changing the radio resource dynamically according to the historical traffic data, current and forecasted traffic, and presented rules described in traffic profiles. The traffic pattern in individual location, area or district can be analyzed and forecasted by data mining the history data, as well as performing real time traffic analysis. This process leads to the action that the planned radio resource can be changed to obtain the optimized composite of voice and data capacities for customers who are using both the voice and data services. Both of the optimized voice and data services are fulfilled by the requirement of a certain “Grade of service” and a pre-defined threshold value.

Figure 3 shows data mining of the GSM network statistics is one of the major steps in the running cycle of RRMS, for defining the best fitted Gos design value in the GSM network. In the second part, the Traffic Policy (TP), traffic policy rule is defined by the radio capacity planner who also referenced to the data mining findings. The defined threshold value should be observed during the radio resource rearrangement in the process. Traffic profile analysis (TPA) is to retrieve and calculate current radio network capacity such as the number of supported users, and the current customer usage. The data calculated from traffic profile analysis is then passed to the Traffic channel reconfiguration (TCR) process for proceeding with the updates on the resource allocations. In the TCR, the optimized Gos, number of Tch & Sdcch, traffic capacity etc., are calculated and the change effect to the GSM network capacity is evaluated. The detailed flow of the whole RRMS operation can be found in [5].

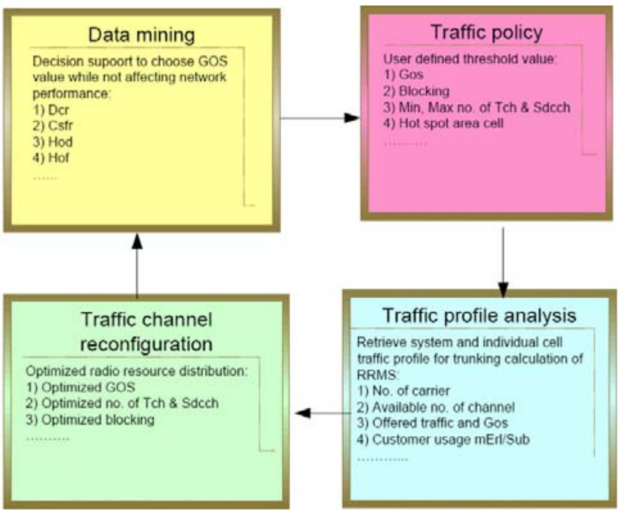


Figure 3. Radio resource management server running cycle

B. *Data Mining of GSM Network Data*

The purpose of data mining in RRMS is to discover the facts and relationship between GSM network performance statistics, and then to provide decision support in choosing an optimized Gos value as a threshold in the Traffic policy. Traditionally, the number of the required voice radio channel of a cell is calculated by Gos and the forecasted peak hour usage. The number of signaling channels is only planned with a default value, which may of course deviate from the actual usage. In the overall RRMS process, the outcome is to optimize and balance the resource arrangement between the number of the planned radio channels for voice and data services. As more value-added services are being supported, including mobile commerce, the demand of data increases as well as fluctuates over different times and over different locations. Also, the new generation network 3G and 4G will also occupy the usage of GSM network since there is internal roaming from 3G/4G to 2G network, and vice-versa due to coverage issue. Both value-added services and 3G/4G traffic increase the loading in the part of the data resource. If the Gos value is set too high, there will be a chance in over estimating the radio resource demand, and the utilization factor will be low. If Gos value is under estimated, inadequate radio resource will lead to network congestion, blocking, call failure and dropped, etc. The number of radio channels available not only affects the probability of the customer making a successful call, it also affects the communication network quality such as call drop, handover failure and drop, or voice quality when a customer call handovers from the current cell to the neighbor cell. When a cell experiences insufficient radio resource for the cells to hand in, the handover failure of the original call will occur. The worst case is if the handover is triggered by coverage or quality problem (coherent channel or adjacent channel interference). When the call cannot handover to neighbor cell instantaneously, the call will have a chance to drop. Call drop is the most undesirable case in the view of a customer.

A designed Gos value is directly related to the number of radio channels being offered, and the radio resource will in turn affect the network performance. Data mining techniques are used to investigate the relationship between each factor, and especially to study the property of the measured Gos in relation to the other key factors which influence the network performance. Once the data are mined, a number of rules will be generated, and they will be used as references for maintaining the GSM network performance as described in the RRM server running cycle. Should the Gos value falls below a certain value, the other key factors such as the call setup successful rate or the call drop rate will also fall within a range. A collection of different statistics, as performance indicators are referred upon for monitoring the network performance. The following statistic measures are chosen as the key factors that represent the overall network performance being monitored at cell level. Some selected key factors and the description of their respective data fields are shown in the table below.

The data to be mined are collected in hourly basis in cell level and daily in district level. In particular, attention was paid to some unusual patterns from mining the data which resulted in abnormal network performance and special utilization of radio resource. The data mining method used here is Association rule for classification [6]. The Association Rule for classification is processed with a minimal confidence value and minimal support value. These minimal values can be chosen corresponding to the certainty and accuracy we want from the resultant rules. It is necessary to mine with association rule with a specific target or class attribute, because in RRMS rules are studied with respective to each factor of the network performance. For several critical attributes, we want to guarantee certain performance levels; for instance, 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) with predefined threshold value. In the following, the Association rules for classification of the target attributes in an example cell are shown:

TABLE I.
KEY DATA FIELDS USED IN DATA MINING

Attribute	Description	Usage
Dcr	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. Statistics may be calculated at the cell, BSS, or network level.	QS, FF, SR
Csfr	Measures the percentage of calls that unsuccessfully access a TCH when service is request for regular calls, emergency calls, SMS procedures, page responses, and calls that are reestablished. Calls which do not require TCH assess are excluded	QS, FF, SA
Hand in	Provide the total number of handover calls from Inter BSS and Intra cell	QS, RA
Busy Tch mean	Records the mean number of TCHs allocated during an interval. This is a weighted distribution statistics and will produce a mean a value indicating the average number of TCHs in use during the interval. It is an indication of the average capacity for additional traffic in the cell. This statistics includes TCHs used as a Dedicated Control Channel (DCCH) in immediate mode.	QS, RF, NP, O
Total Calls	Counts the total number of calls originated for each cell on the BSS	NP
Handover failure	Percentage of attempted handovers that fail with the MS recovering to the source cell	QS, FF
TCH Gos and TCH blocking	Percentage of total number of unsuccessful allocations of TCH	QS, FF
IC Erlang	This statistic provide the ratio of Intra Cell Handover Per Erlang	QS, RF
Ioi	Indicates the maximum, minimum, and mean interference levels experienced on idle channels as determined by the BSS	QS, RF, NP, FF
Ber	Indicates the Bit Error Rate (Ber) for active channels. The bit error rate value corresponds to the RXQUAL downlink measurement reports. BER statistic measurements are reported only when the channel is active	QS, RF, FF

Legends: Usage: FF – Fault finding, SR – Service retainability, SA – Service accessibility, QS – Quality of service, RA – Resource allocation, RF – RF loss, NP – Network Planning, O – Optimization

The relevant key statistical data are collected and processed by OMC. Data mining analyst 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.

As a matter of data preprocessing prior to data mining, the raw data will have to be computed into key statistics like those in Table 1 via some predefined formula. The key statistics which are numeric data by nature will be discretized into ranges of different bins for Association Rule mining.

IV. RESULTS OF MINING GSM NETWORK DATA

Various data mining methods are applied in order to discover interesting patterns at RRMS, they include spatial visualization, clustering, association rule and association rule for classification. Spatial visualization is particularly useful to monitor base stations performance and to detect if any abnormal event occurs. The other mining tools assist in searching for useful and interesting patterns, trends, and rule information etc. All these knowledge obtained can help in dimensioning the network loading and maintaining the network performance. Castaneda Data Mining Tools are used in our experiments (Cf. Castaneda DMS, www.girgese.com). The findings are very comprehensive. However, some selected sample results are show here for illustrating the feasibility of the data mining tool. The data used in the experiments that are based on real data, have been modified from their original distributions for privacy and confidentiality reasons.

A. Spatial Visualization

Spatial visualization [7] in this project is mainly about displaying the statues of the network usage in terms of base stations performance and conditions via visual colors and shapes, spatially and temporally over a regional map. Since the radio coverage of GSM networks is composed of a number of base stations in different locations, it is necessary to monitor the performance of each cell site on a regular basis. In this project, we have mapped parameters such as Traffic, Tch Gos, Csfr and Dcr by geographical means, i.e., its distribution among the proximity of Macao. Optionally, we can assign other attributes for displaying the results on a geographical map. For traffic (busy_tch_mean), it shows the distribution of traffic density in different areas. For the Tch Gos, it shows which area has a higher Tch gos, and so forth.

These visualization results can be used for network traffic dimensioning and resources management [8], and it can easily alert traffic problems and aid in shifting some high traffic areas to other areas (from Red to Green) in order to balance traffic loading. For monitoring Csfr and Dcr, those areas with red or orange spots are hotspots of high call setup failure and drop call rate. With spontaneous visual feedback, network operators can relocate resources to ensure well performing base stations.

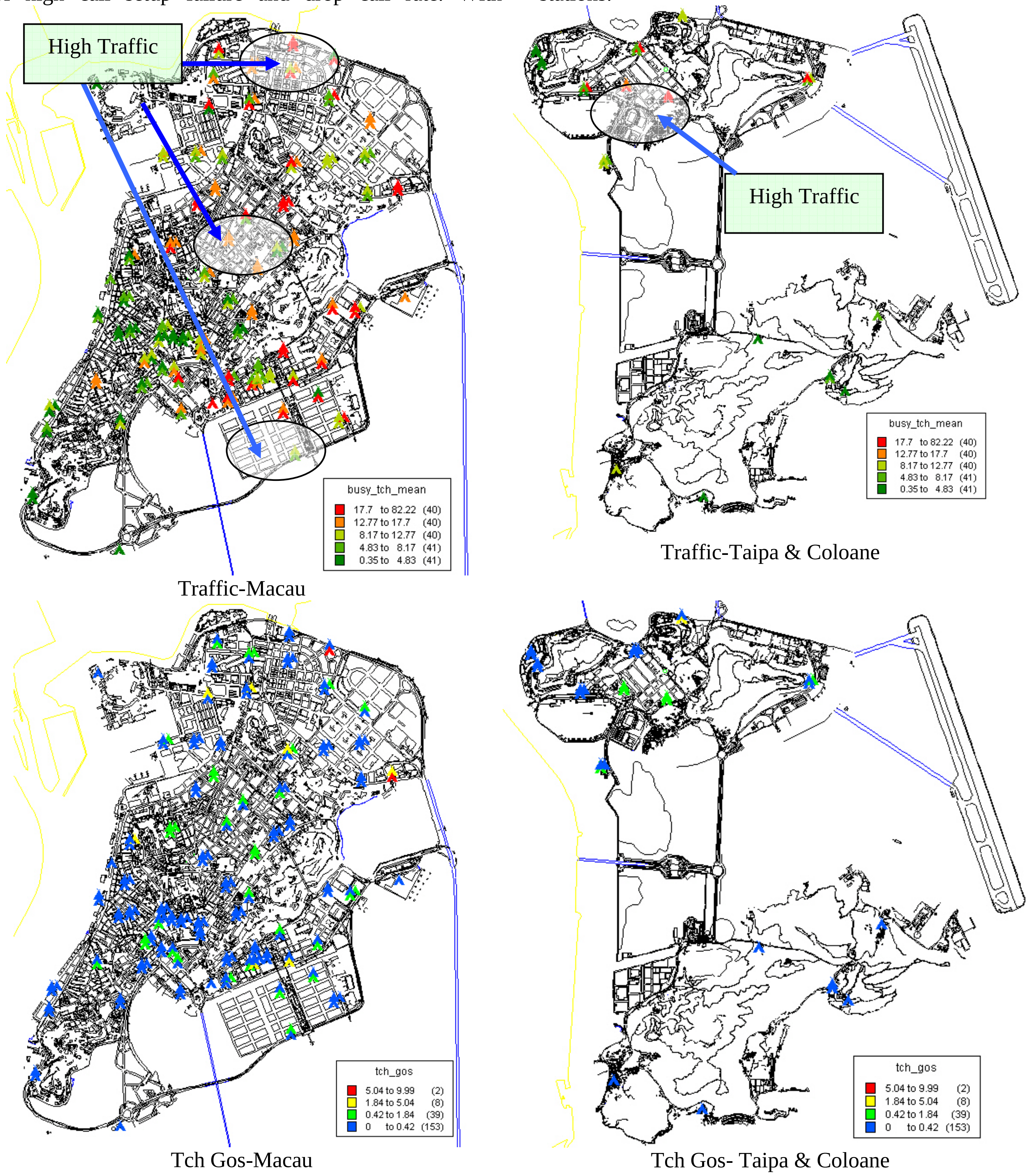


Figure 4. Spatial visualization of GSM network performance over Macau – examples of Traffic and Tch Gos

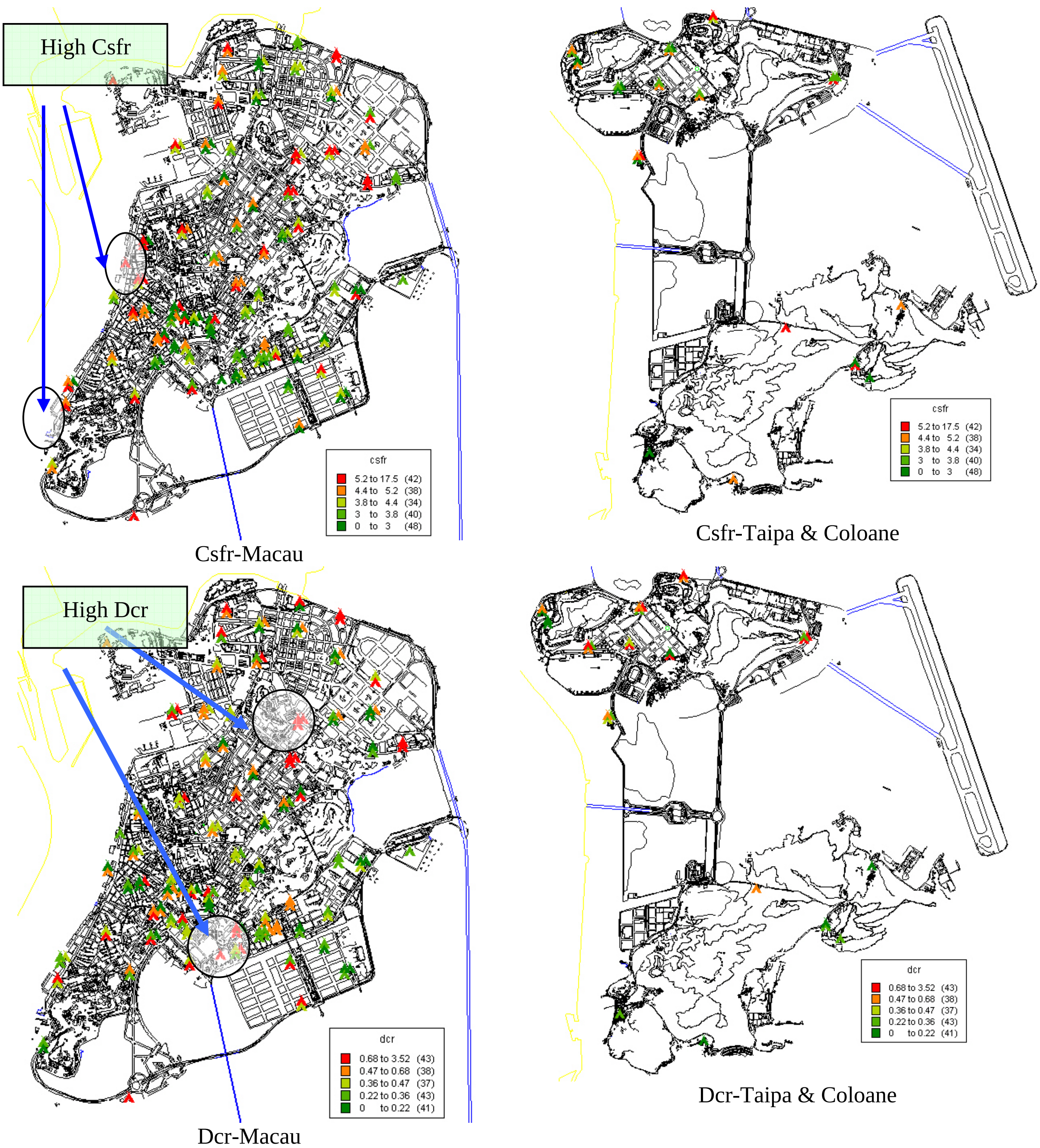


Figure 5. Spatial visualization of GSM network performance over Macau – examples of Csfr and Dcr.

B. Traffic analysis by classification

By weekday: We start by aggregating weekly data on weekday basis and try to find the distribution of several parameters over each day of week. Castaneda classifies each parameter into 5 groups which simplify the visualization process. We try to analyze traffic, Tch blocking, Sdcch blocking over each day of week in order to gain deeper knowledge of high traffic and blocking dates. By gaining access to this type of information, we can better allocate our resources and equipment in keys dates such as days of week with high traffic.

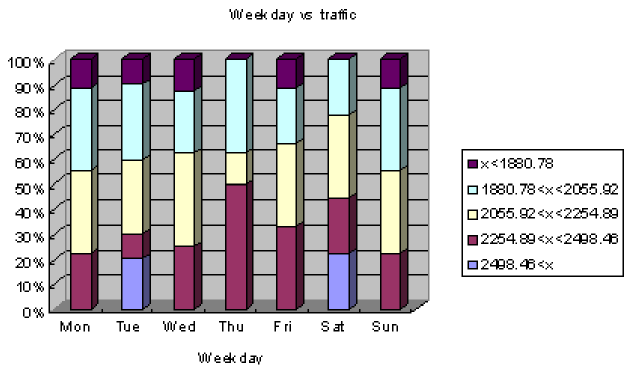


Figure 6. Traffic distribution by weekday.

By District: Similarly, we try to aggregate traffic, TCH blocking and SDCCH blocking in order to find regions with high traffic and blocking rate.

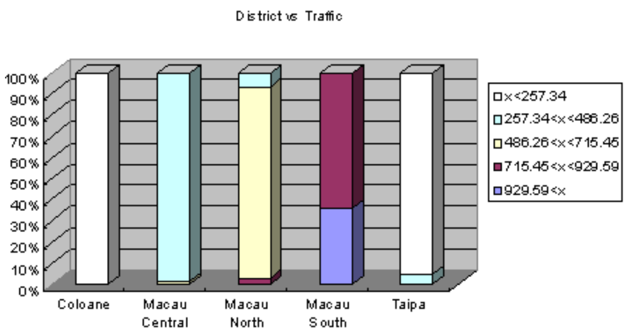


Figure 7. Traffic distribution by district.

Dimensioning by Tch Gos: Grade of service (Gos) is defined as the probability that a subscriber cannot make a call successfully. Lower Gos means better quality of service. The following figure shows how busy peak traffic hours can affect the Gos.

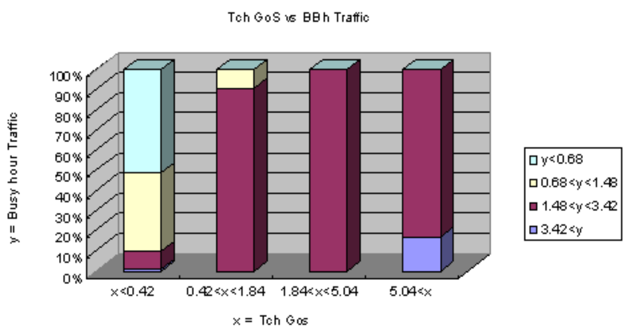


Figure 8. Tch GoS by BBh Traffic.

When the pear hour traffic fulfills the condition $y < 0.68$ or $y < 1.48$, about 90% of GOS is under 1.84. However, if BBh traffic is y higher than 3.42, then Gos goes up to 5.04. It means that for peak hour traffics, the quality of service is lowered.

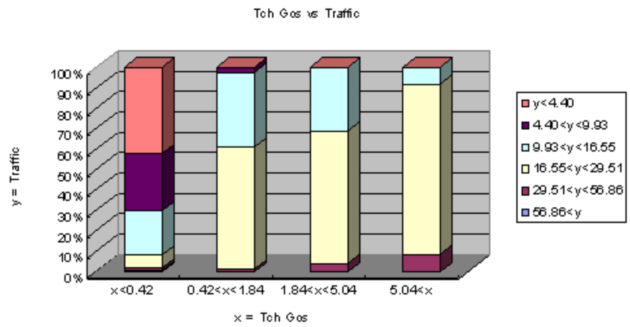


Figure 9. Tch GoS by Traffic.

Compared to Figure 8, Figure 9 gives a general idea on how to dimension Gos with daily traffics. It is found that when traffic is lower than 16.55, the Gos can be maintained at a level smaller than 0.42.

The classification experiments can consider changes of Gos by the number of Daily Total Calls, when total call is larger than 1074, then Gos is most likely to be greater than 0.42 and smaller than 1.84. Some phenomenon can be observed by changing the attribute to Daily sum of Total Call and call handovers from other cells. When sum is $3376 < y < 5290$ or $y > 5290$, then the Gos is most likely greater than 5.04. It hints us to tune the parameter in order to prevent handovers form neighbor concentrated into one serving cell.

By Hour Distribution: We try to analyze peak hour traffic by aggregating hourly base data versus traffic. As shown in Figure 10, starting at 12p.m., traffic is increasing with blue bar to a peak density at 6p.m., and then decreases gradually till 12 a.m. At 12 a.m., traffic decreases continuously to 8a.m. and starts to increase again.

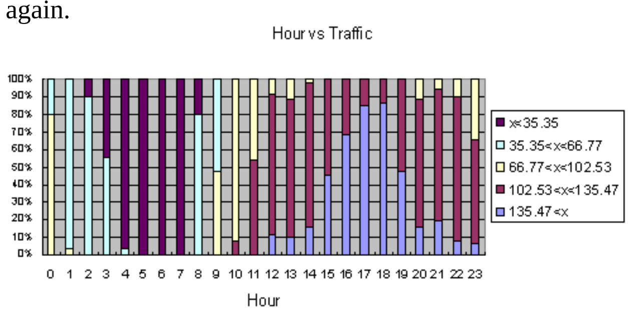
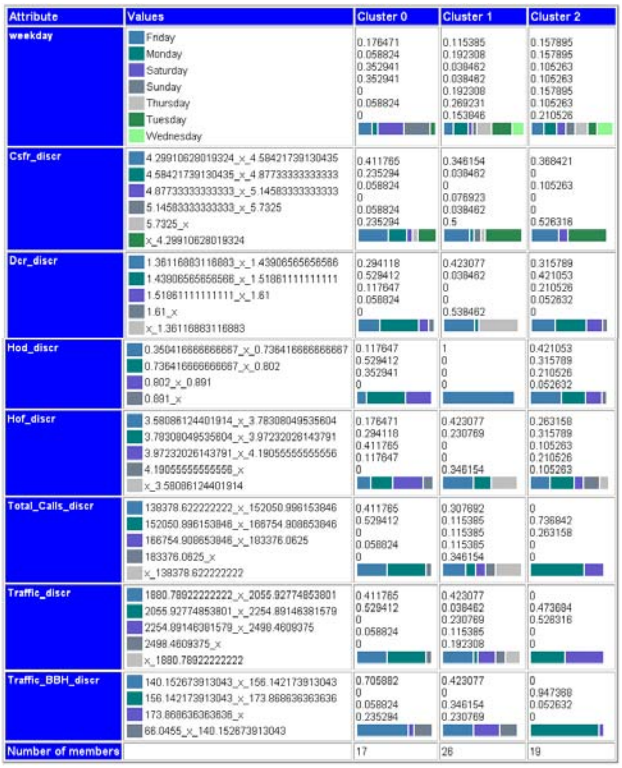


Figure 10. Traffic by Hour.

Similar experiments on hourly traffic can be done for Tch and Sdcch blocking. Starting from 10a.m., our results indicate that Tch blocking is increasing with a red bar to a peak density at hour 6p.m. and decreases gradually till hour 6a.m. again. At hour 7a.m. the blocking decreases continuously until hour 8a.m. and starts to increase again. For Sdcch blocking distribution by hour, blocking is always under a satisfactory rate of 2.629. Importantly by understanding the traffic distribution, Tch blocking during peak hours can be prevented.

B. Traffic analysis by clustering

Using clustering [9], some behavioral insight of the traffic data could be observed especially when there are too many input attributes. This can clarify the healthy status of the system at different levels, such as: system daily, system hourly, and district daily and district hourly levels. By analyzing the following 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 which cells within this range, then to locate the problematic cells that caused a high Dcr.



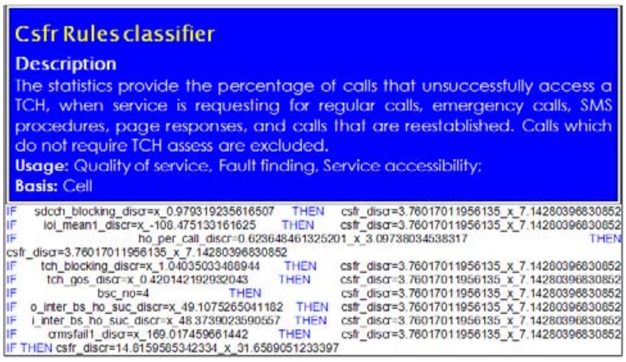
Attribute	Values	Cluster 0	Cluster 1	Cluster 2
weekday	Friday Monday Saturday Sunday Thursday Tuesday Wednesday	0.176471 0.058824 0.352941 0.352941 0 0.058824 0	0.115385 0.192308 0.038462 0.038462 0.192308 0.269231 0.153846	0.157895 0.157895 0.105263 0.105263 0.157895 0.105263 0.210526
Csfr_discr	4.29910628019324_x_4.58421739130435 4.58421739130435_x_4.87733333333333 4.87733333333333_x_5.14583333333333 5.14583333333333_x_5.7325 5.7325_x x_4.29910628019324	0.411765 0.235294 0.058824 0 0.058824 0.235294	0.346154 0.038462 0 0.076923 0.038462 0.5	0.368421 0 0.105263 0 0 0.526316
Dcr_discr	1.36116883116883_x_1.43906565656566 1.43906565656566_x_1.51861111111111 1.51861111111111_x_1.61 1.61_x x_1.36116883116883	0.294118 0.529412 0.117647 0.058824 0	0.423077 0.038462 0 0 0.538462	0.315789 0.421053 0.210526 0.052632 0
Hod_discr	0.350416666666667_x_0.736416666666667 0.736416666666667_x_0.802 0.802_x_0.891 0.891_x	0.117647 0.529412 0.352941 0	1 0 0 0	0.421053 0.315789 0.210526 0.052632
Hof_discr	3.58086124401914_x_3.78308049535604 3.78308049535604_x_3.97232026143791 3.97232026143791_x_4.19055555555556 4.19055555555556_x x_3.58086124401914	0.176471 0.294118 0.411765 0.117647 0	0.423077 0.230769 0 0 0.346154	0.263158 0.315789 0.105263 0.210526 0.105263
Total_Calls_discr	138378.622222222_x_152050.996153846 152050.996153846_x_166754.908653846 166754.908653846_x_183376.0625 183376.0625_x x_138378.622222222	0.411765 0.529412 0 0.058824 0	0.307692 0.115385 0.115385 0.115385 0.346154	0 0.736842 0.263158 0 0
Traffic_discr	1880.78922222222_x_2055.92774853801 2055.92774853801_x_2254.89146381579 2254.89146381579_x_2498.4609375 2498.4609375_x x_1880.78922222222	0.411765 0.529412 0 0.058824 0	0.423077 0.038462 0.230769 0.115385 0.192308	0 0.473684 0.526316 0 0
Traffic_BBH_discr	140.152673913043_x_156.142173913043 156.142173913043_x_173.868636363636 173.868636363636_x 66.0455_x_140.152673913043	0.705882 0 0.058824 0.235294	0.423077 0 0.346154 0.230769	0 0.947368 0.052632 0
Number of members		17	26	19

Figure 11. Clustering of network performance by day distribution.

C. Traffic analysis by Association Rule for Classification

In this project we applied Association Rule for Classification [10] to find useful rules for controlling the network performance by maintaining the essential attributes, such as 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. Instead of being exhaustive only Macau North District was used for analysis.

The association rule and the Association Rule for classification are processed with a minimal confidence 0.5 and a minimal support 0.5. However, it is wise to mine with association rule with a target attribute; otherwise it may give overwhelmingly many different combinations of association rules. Using association rule is suitable when the data source does not contain more than 10 columns as a common practice. By using Association Rule for classification, it can give some general rules on how to maintain the network performance. The symbol _x_ represents it is between two values such as $y<x<z$. An important task at RRMS is traffic profile analysis based on current cell site traffic and by considering the mined results.



Csfr Rules classifier

Description

The statistics provide the percentage of calls that unsuccessfully access a TCH, when service is requesting for regular calls, emergency calls, SMS procedures, page responses, and calls that are reestablished. Calls which do not require TCH assess are excluded.

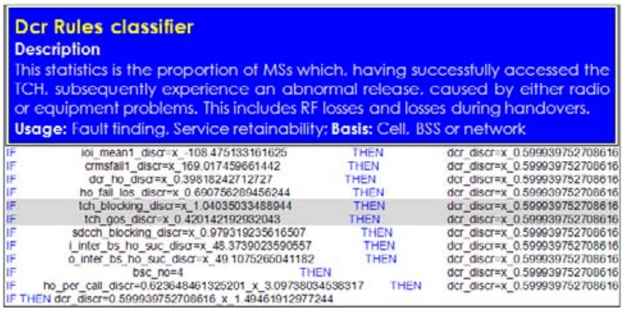
Usage: Quality of service, Fault finding, Service accessibility;

Basis: Cell

IF sdcch_blocking_discr=x_0.979319235616507 THEN csfr_discr=3.76017011956135_x_7.14280396830852
IF ioi_mean1_discr=x_-108.475133161625 THEN csfr_discr=3.76017011956135_x_7.14280396830852
IF ho_per_call_discr=0.623648461325201_x_3.09738034538317 THEN csfr_discr=3.76017011956135_x_7.14280396830852
IF tch_blocking_discr=x_1.04035033488944 THEN csfr_discr=3.76017011956135_x_7.14280396830852
IF tch_gos_discr=x_0.420142192932043 THEN csfr_discr=3.76017011956135_x_7.14280396830852
IF bsc_no=4 THEN csfr_discr=3.76017011956135_x_7.14280396830852
IF o_inter_bs_ho_suc_discr=x_49.1075265041182 THEN csfr_discr=3.76017011956135_x_7.14280396830852
IF i_inter_bs_ho_suc_discr=x_48.3739023590557 THEN csfr_discr=3.76017011956135_x_7.14280396830852
IF crmsfail1_discr=x_169.017459661442 THEN csfr_discr=3.76017011956135_x_7.14280396830852
IF THEN csfr_discr=14.8159585342334_x_31.6589051233397

Figure 12. Csfr Rules Classifier.

From the technical perspective of radio network, 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 solve 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 degradation 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). Moreover, Gos (Tch Gos) for each cell is related to its current number of Tch and Traffic as shown by the association rules in Figure 12.



Dcr Rules classifier

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 sdcch_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 13. Dcr Rules 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 handovers also will cause BTS overloaded 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.

Dcr Ho
Description
This statistics is the proportion of MSs which, having successfully accessed the TCH, subsequently experience an abnormal release during handover caused by either radio or equipment problems. This includes losses during handovers.
Usage: Fault finding, Service retainability;
Basis: Cell, BSS or network

IF	ioi_mean1_discr=x_-108.475133161625	THEN	dcr_ho_discr=x_0.39818242712727
IF	crmsfail1_discr=x_169.017459661442	THEN	dcr_ho_discr=x_0.39818242712727
IF	ho_fail_los_discr=x_0.690756289456244	THEN	dcr_ho_discr=x_0.39818242712727
IF	tch_blocking_discr=x_1.04035033488944	THEN	dcr_ho_discr=x_0.39818242712727
IF	tch_gos_discr=x_0.420142192932043	THEN	dcr_ho_discr=x_0.39818242712727
IF	sdcch_blocking_discr=x_0.979319235616507	THEN	dcr_ho_discr=x_0.39818242712727
IF	o_inter_bs_ho_suc_discr=x_49.1075265041182	THEN	dcr_ho_discr=x_0.39818242712727
IF	bsc_no=4	THEN	dcr_ho_discr=x_0.39818242712727
IF	ho_per_call_discr=0.623648461325201_x_3.09738034538317	THEN	dcr_ho_discr=x_0.39818242712727
IF	i_inter_bs_ho_suc_discr=x_48.3739023590557	THEN	dcr_ho_discr=x_0.39818242712727
IF THEN dcr_ho_discr=0.39818242712727_x_1.1976807932936			

Figure 14. Dcr Ho Rules Classifier.

The most useful rules for Ho Fail los & Ho Fail and Ret are their relations to Ioi_mean, Sdcch blocking, Ho_per_call. The relations 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 interfered by others by keeping Sdcch blocking in a low level. This can ensure that a cell has sufficient 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 problems can be cleared. It is not usual for a phone call to handover for more than three times when it is serving a stable and well-covered area.

Ho Fail Lost Rules classifier
Description
This statistics is the percentage of attempted handovers that result in lost connections.
Usage: Quality of Service, Fault finding

IF	sdcch_blocking_discr=x_0.979319235616507	THEN	ho_fail_los_discr=x_0.690756289456244
IF	bsc_no=4	THEN	ho_fail_los_discr=x_0.690756289456244
IF	dcr_ho_discr=x_0.39818242712727	THEN	ho_fail_los_discr=x_0.690756289456244
IF	o_inter_bs_ho_suc_discr=x_49.1075265041182	THEN	ho_fail_los_discr=x_0.690756289456244
IF	i_inter_bs_ho_suc_discr=x_48.3739023590557	THEN	ho_fail_los_discr=x_0.690756289456244
IF	tch_blocking_discr=x_1.04035033488944	THEN	ho_fail_los_discr=x_0.690756289456244
IF	tch_gos_discr=x_0.420142192932043	THEN	ho_fail_los_discr=x_0.690756289456244
IF	ioi_mean1_discr=x_-108.475133161625	THEN	ho_fail_los_discr=x_0.690756289456244
IF	ho_per_call_discr=0.623648461325201_x_3.09738034538317	THEN	ho_fail_los_discr=x_0.690756289456244
IF	crmsfail1_discr=x_169.017459661442	THEN	ho_fail_los_discr=x_0.690756289456244
IF THEN ho_fail_los_discr=0.690756289456244_x_2.08135432069557			

Figure 15. Ho Fail Lost Rules Classifier.

Ho Fail and Ret Rules classifier
Description
This statistics is the percentage of attempted handovers that fail with the MS recovering to the source cell
Usage: Quality of Service, Fault Finding
Basis: Cell

IF	ioi_mean1_discr=x_-108.475133161625	THEN	ho_fail_ret_discr=x_3.45155113380181
IF	sdcch_blocking_discr=x_0.979319235616507	THEN	ho_fail_ret_discr=x_3.45155113380181
IF	crmsfail1_discr=x_169.017459661442	THEN	ho_fail_ret_discr=x_3.45155113380181
IF	ho_per_call_discr=0.623648461325201_x_3.09738034538317	THEN	ho_fail_ret_discr=x_3.45155113380181
IF	bsc_no=4	THEN	ho_fail_ret_discr=x_3.45155113380181
IF	tch_gos_discr=x_0.420142192932043	THEN	ho_fail_ret_discr=x_3.45155113380181
IF	tch_blocking_discr=x_1.04035033488944	THEN	ho_fail_ret_discr=x_3.45155113380181
IF THEN ho_fail_ret_discr=18.3321747289211_x_34.8190416088766			

Figure 16. Ho Fail and Ret Rules Classifier.

CONCLUSION

Traditionally, GSM service providers have related the number of base stations directly with the network quality and performance. They focused on increasing the numbers of stations to improve the quality of service. The cost of ownership and maintenance increases as the number of stations increases proportionally. In current GSM system, there is OMCR (Operations and Maintenance Centre - Radio Part) for monitoring real-time network system performance by checking for any different abnormalities, from base station faults, alarms to power supply problems. Factors like outstanding cell performance, high Dcr (Drop Call Rate) or Tch (Traffic CHannel) blocking will also affect network quality. Instead of monitoring a particular cell, the whole network system is examined, which might be composed by over hundreds to ten thousands of individual cells in our proposed methodology.

Being able to monitor and analyze the operation of a whole network is particularly important because nowadays mobile network will have to support different kind of value added service including commercial mobile payment applications. Resource planning and performance monitoring should be quick, complete and automated as much as possible. Traditionally the network statistics are monitoring by comparing with a system-calculated figure and the irregularities are analyzed manually by a group of technique people. By applying data mining techniques (with support of a good database management system), interesting patterns, rules, or any anomaly plus their relevant data could be located easily. The operator or radio planning engineer can obtain valuable information from it sooner, and be able to react to problems more quickly.

By means of spatial visualization the traffic channel monitoring is easy and remarkably noticeable at a glance. Traffic analysis by classification provides network engineers the usage patterns over each slot of time-band, day-band versus others attributes such as district, cell handover, etc. Several other attributes are also put into analysis; not only the traffic blocking but also other key indicators such as drop call and call setup failure. Clustering segments data groups by their similarity. 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, the engineer can keep the Dcr at a certain acceptable rate by controlling the traffic channel at a certain level.

In summary, with the aid of data mining, the network performance could be more effectively monitored, and from which provides a better configuration and QoS support [11]. The work contributed in this paper laid a foundation for future hierarchical QoS management [11]. This also offers decision support for the radio planning engineer to configure and fine-tune the mobile network. Based on the decision support and the traffic profile analysis and prediction, radio planning engineer can plan the radio resources more effectively, and increase the utilization rate. It can provide a well-planned resource proportion for voice and data services, hence supporting reliably mobile payment applications.

Besides providing healthy and enough resource in the radio side, the monitoring of the core network in the switching side is also very important. In this project, we have also investigated the QoS monitoring parameters of data service at OMCS side. Keeping a network in a good performance at both radio level and cell switching level, is one of the key success factors to provide a reliable and stable value added service.

ACKNOWLEDGMENT

The former MSc. Graduates, Edison Lai, Henry Ho and Antonio Leung, from University of Macau have contributed greatly in the experiments. Their efforts are highly appreciated.

REFERENCES

[1] S. Fong, and E. Lai, "Mobile Mini-payment Scheme Using SMS-Credit", *International Conference of Computational Science and Its Applications, ICCSA 2005*, Lecture Notes in Computer Science, Springer-Verlag, Vol.2., Singapore, pp. 1106-1116, 9-12 May 2005.

[2] H. Abu-Ghazaleh, A.S. Alfa, "Application of Mobility Prediction in Wireless Networks Using Markov Renewal Theory", *IEEE Transactions on Vehicular Technology*, Vol. 59, Iss. 2, pp.788-802, Feburary 2010.

[3] I.R. Chen, N. Verma, "Simulation Study of a Class of Autonomous Host-Centric Mobility Prediction Algorithms for Wireless Cellular and Ad Hoc Networks", *ANSS '03: Proceedings of the 36th annual symposium on Simulation*, IEEE Computer Society, March 2003

[4] A. Leung, S. Fong, E. Lai, "Mining Operational Data for Improving GSM Network Performance", *International Conference on Fuzzy Systems and Knowledge Discovery*: *Computational Intelligence for the E-Age (FSKD 2002)*, Orchid Country Club, Singapore, ISBN 981-04-7520-9, pp.285-289, 18-22 November 2002.

[5] E. Lai, S. Fong, H. Yang, "Supporting Mobile Payment QOS by Data Mining GSM Network Traffic", *The 10th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2008)*, Linz, Austria, ACM Press, ISBN:978-1-60558-349-5, pp.279-285, 24-26 November 2008.

[6] G. Chen, H. Liu, L. Yua, Q. Wei and X. Zhang, "A New Approach to Classification Based on Association Rule Mining", *Decision Support Systems*, Volume 42, Issue 2, Elsevier, pp.674-689, November 2006.

[7] A. Vaidyanathan, M. Billinghurst, H. Sirisena, "Characterizing and Visualizing Mobile Networks", NZCSRSC 2008, Christchurch, New Zealand, pp.165- 169, April 2008.

[8] R.T. Valadas, "Dimensioning and Resource Management of ATM Networks", *7th IFIP/ICCC Conference on Information Networks and Data Communications, (INDC'98)*, pp.209-220, 1998.

[9] A. Hushyar, "Network Traffic Clustering and Geographic Visualization", Master's Theses, *San Jose State University*, USA, Paper 3695, pp.1-42, 2009.

[10] U. Chaudhary, I. Papapanagiotou, M. Devetsikiotis, "Flow Classification Using Clustering And Association Rule Mining", The International Workshop on Computer-Aided Modeling Analysis and Design of Communication Links and Networks (IEEE CAMAD 2010), December 3-4, 2010.

[11] S. Fong, "Adaptive QoS Resource Management by Using Hierarchical Distributed Classification for Future Generation Networks", The Third International Conference on Computer Networks & Communications (CoNeCo-2011), Mesa, Ankara, Turkey, June 26-29, 2011

Simon Fong graduated from La Trobe University, Australia, with a 1st Class Honours BEng. Computer Systems degree and a PhD. Computer Science degree in 1993 and 1998 respectively. Simon is now working as an Assistant Professor at the Computer and Information Science Department of the University of Macau. He is also one of the founding members of the Data Analytics and Collaborative Computing Research Group in the Faculty of Science and Technology. Prior to joining the University of Macau, he worked as an Assistant Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. Before his academic career, Simon took up various managerial and technical posts, such as systems engineer, senior IT consultant and e-commerce director in Melbourne, Hong Kong and Singapore. Some companies that he worked include Hong Kong Telecom, Singapore Network Services, AES Pro-Data Australia and United Oversea Bank, Singapore. Dr. Fong has published over 120 peer-referred international conference and journal papers, mostly in the areas of E-Commerce and Data-mining.