

Supporting Mobile Payment QOS by Data Mining GSM Network Traffic

Edison Lai
Faculty of Science and Technology
University of Macau
Macao SAR, China

Simon Fong
Faculty of Science and Technology
University of Macau
Macao SAR, China
+853 3974473
ccfong@umac.mo

Yang Hang
Faculty of Science and Technology
University of Macau
Macao SAR, China
+853 66874953
ma76562@umac.mo

ABSTRACT

In mobile commerce, short-message-service (SMS) is an important technique for delivering payment instruction. A payment model "SMS Credit" was proposed earlier [1]. Such payment service or similar relies on the transmission of SMS; it is needed to reduce the occurrence of packet losses and delay, to improve the quality of packet transmission services (QOS) in the network. This paper discusses how the payment service operates in a configurable radio resource environment via data mining. A Radio Resource Management and Prediction Server equipped with data mining algorithms will optimize the radio resources for both voice and data services in order to provide an optimized QOS. Specifically, data mining techniques are applied to define traffic policy and to calculate optimization result through traffic profile analysis.

Categories and Subject Descriptors

H.2.8 [Database Applications Data mining]: Data Mining

General Terms

Algorithms, Management, Measurement, Documentation, Performance, Design, Economics, Reliability, Experimentation, Security, Human Factors, Standardization, Languages, Theory, Legal Aspects, Verification.

Keywords

GSM Network, Grade of Service, Quality of Service, Data-Mining.

1. INTRODUCTION

Nowadays, mobile networks need to support a high density of user population and multimedia services. In particular, mobile commerce that leverages on SMS has a stringent set of performance requirements on the delivery of messages, such as

loss and latency [2].

Given a limited capacity of resources by the fixed hardware infrastructure, mobile network is statistically allowed to relocate the radio resource based on the predication of the radio resource or traffic profile. Some problem 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 hence has become an important topic in mobile network management [3]. In this paper, we show that data mining algorithms applied for predicting and managing the radio resource movement of cell. Cell usage and performance are mined from the historical data, and then the prediction rules are extracted from these patterns. That follows radio resource management and reconfiguration is accomplished by using these rules.

The radio resource management in the mobile network covers the methods for storing and updating the usage and the traffic profile of the mobile users who are served by the system. Radio resource prediction can be defined as the prediction of a cell next usage movement, where the cell is serving for different mobile user traveling in and out [4]. The predicated movement can then be used to increase the efficiency and utilization of mobile network.

By using the predicted and relocation scheme, the system can effectively allocate resources in the cell neighborhood of a mobile user. Effective relocation of resources to mobile users would improve resource utilization and reduce the latency in accessing the resources. Accurate prediction and relocation is also crucial in processing the queries of mobile payment service, especially in the m-payment model of SMS Credit [1]. When a user submits a payment service request or query, the response or processing liability will depend on the current cell resource at which the user is located. With effective predication and relocation, it is also possible to transfer the queries of the payment service from one cell to the other cell which has sufficient resource.

The paper is organized in this way. Section 2 introduces the disadvantages of the current GOS measure to mobile payment service. It shows the deficiency mainly come from the manual setting of a fixed GOS that results under- or over- utilization of radio resources. In section 3 we proposed a data-mining solution

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

iiWAS2008, November 24–26, 2008, Linz, Austria.

(c) 2008 ACM 978-1-60558-349-5/08/0011 \$5.00.

for resolving the problems. Section 4 showcases some results of mining GSM network data. A conclusion is drawn at the end.

2. DISADVANTAGE OF THE CURRENT GRADE OF SERVICE MEASURE

Mobile networks rely on trunking to accommodate a large number of users in a limited radio spectrum. The concept of trunking allows a large number of users to share a relatively small number of channels in a cell by providing access to each user, on demand, from a pool of available channels [5]. In this trunked radio system, each user is allocated a channel on a per call basis, and upon termination of the call, the previously occupied channel is immediately returned to the pool of available channels.

Trunking exploits the statistical behavior of users so that a fixed number of channels or circuits may accommodate a large, random user's community. The same principle is used in designing mobile networks.

The grade of service (GOS) is a measure of the ability of a user to access a trunked system during the busiest hour [6]. The busy hour is based upon customer demand at the busiest hour during a day, a week, a month, or a year. The grade of service is a benchmark used to define the desired performance of a particular trunked system by specifying a desired likelihood of user obtaining channel access given a specific number of channels available in the system. It is the job of a radio planning engineer or designer to estimate the maximum required capacity and to allocate the proper number of channels in order to meet the GOS requirement.

The radio resource planning methods are mostly focused on the problem of estimating the GOS, which is concerned with the blocking rate by a mobile user who successfully makes a service request. The GOS value only considers whenever a mobile user can make a call or service request in a cell. Compared with the neighbor cell usage, radio channel and data channel resource distribution, it lacks the flexibility and the overall combination effect from a larger view. This type of radio resource planning has some deficiencies, which are explained as follows.

In the general GSM network design, radio frequency planners only use a constant GOS value such as 2% or 5% in manually planning the capacity of radio networks. The values are chosen from the working experience of some senior planners. This is however prone to problems of under- or over-allocation of resources intermittently if not frequently.

In this paper, we propose to use data mining methods to find out what the best GOS value is in different situations, as a dynamic variable for planning the radio network capacity. Using data mining, we can find a best fitted GOS, and the related number of traffic channels (TCH) value which can maximize the radio capacity and utilization while keeping the network in a healthy condition. The healthy condition means keeping in check of the Call setup failure rate (Csfr), Drop call rate (Dcr), Handover drop rate (Hod), etc., in a worst condition. In choosing the appropriate GOS value, there are two types of trunked systems which are commonly used. One type has queuing and the other type offers no queuing for call requests. That is, for every user who requests a service, it is assumed there is no setup time and the user is given immediate access when resource is available. Or

else he is blocked without access and is free to try again later. This type of trunking is called blocked call cleared and assumes that the calls arrive as determined by a Poisson distribution. Furthermore, it is assumed that there are an infinite number of users as well as the following: (a) there are memory-less arrivals of requests, implying that all users, including blocked users, may request a channel at any time.; (b) the probability of a user occupying a channel is exponentially distributed, so that longer calls are less likely to occur as described by an exponential distribution; and (c) there are a finite number of channels available in the trunking pool. This is known as a M/M/m queue, and leads to the derivation of the Erlang B formula (also known as the blocked calls cleared formula) [7]. The Erlang B formula determines the probability that a call is blocked and it is a measure of the GOS for a trunked system which provides no queuing for blocked calls.

In the mobile payment model that operates on GSM network, Blocked call cleared is generally assumed and investigated in the traffic optimization.

In this paper, we aim to set a GOS using data mining, while not decreasing the offered traffic, but offering the same GOS or predefined threshold GOS value. Thus the only variable is the change of resource allocation between voice and data, by altering the number of traffic channels (TCH) and signaling channels (SDCCH).

3. PROPOSED DATA-MINING SOLUTION

To overcome the deficiencies, we have developed an effective radio resource management and prediction system with data-mining capability as shown in Figure 1. It is called Radio Resource Management and Prediction Server (RRMPS). The goal of the RRMPS is to optimize the traffic loads for different radio channel configurations in GSM cell sites. The challenge now is to determine how many channels or resources are available in each site/system in the RRMPS.

As shown in Figure 2, it shows the eight stages for providing data mining support to mobile payment service.

- 1) Operation and Maintenance Center-Radio (OMCR) and Operation and Maintenance Center-Switching (OMCS) collect the raw data from different network elements and build up the database in terms of the collected raw data;
- 2) In the database, it needs to perform some pre-defined calculation to change the raw data into key statistics;
- 3) Before applying various data mining techniques, the continuation of the input data or attributes is discretized into different ranges by k-mean;
- 4) By association rule, visualization map, and clustering, the output is the decision support for RRMPS;
- 5) The RRMPS evaluates the possibilities and calculates the reconfiguration combinations according to the traffic profile [8], analysis and the decision support induced by data mining. Operator will modify the cell site voice and data channel allocation and distribution based on the information provided from it;
- 6) QOS monitoring of SMS is conducted at the short message service center (SMSC), such as the CPU loading or SMS Mobile terminated/Mobile Originated delivery successful rate. It will

display abnormal event or alarm which gives the threshold value for engineer to follow

7) SMS credit payment server is running as a value added service in mobile network and providing the mobile payment service. This server can communicate with the SMSC.

8) Customers use mobile phones to perform SMS credit based payment at authorized merchants.

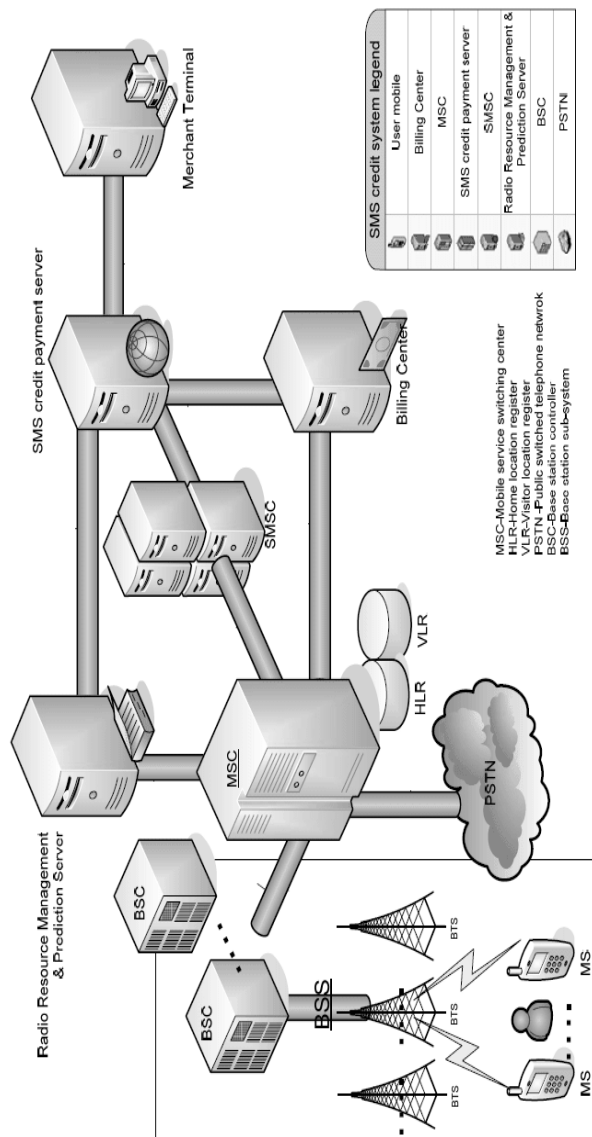


Figure 1 SMS credit payment model with RRMPs

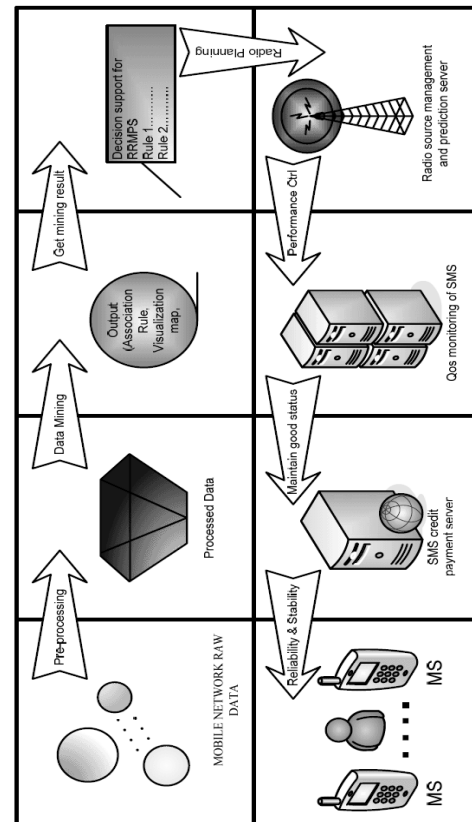


Figure 2 Frame work diagram of data mining for supporting payment service

3.1 Radio Resource Management and Prediction Server (RRMPs)

In order to optimize and plan the radio resource for both voice and data service in GSM network, the concept of changing the radio resource dynamically, which is according to the historical traffic data, current and forecasted traffic, presented rule or policy, is proposed. The traffic pattern in individual location, area or district can be analyzed and forecasted or projected due to historical data, and combining real time traffic analysis. This leads to the action that the planned radio resource can be changed to obtain the optimize voice and data capacity for customer while using the voice and data services. Both the optimized voice and data services are fulfilled by a suitable GOS and a pre-define GOS threshold value.

In Figure 3, the RRMPs running cycle, data mining of the GSM network statistics is one of the major parts for defining the best fitted GOS design value. The Traffic Policy (TP) rule is defined by radio capacity planner, and according to the results found by data mining. During re-arrangement, the defined threshold value should be obeyed during the radio resource rearrangement. Traffic profile analysis (TPA) is to retrieve and calculate current radio network capacity such as the number of supported users and the usage. The data calculated from traffic profile analysis is then passed to the Traffic channel reconfiguration (TCR) process. In the TCR, the optimized GOS,

number of TCH & SDCCH, traffic capacity etc., are calculated. As well as the change effect of the GSM network capacity is evaluated.

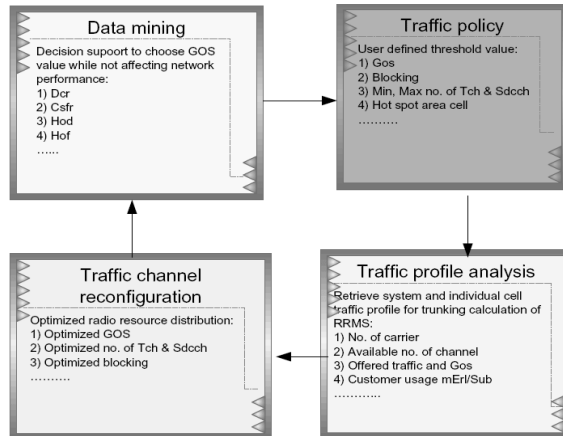


Figure 3 Radio Resource Management and Prediction Server running cycle

3.2 Mining the GSM Network Data

The purpose of data mining in RRMPs is to discover the facts and relationship between GSM network performance data, and then to generate a decision support in choosing a GOS value for defining the threshold value 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 channel is only planned with the default value, which sometimes may not meet the requirements of the actual usage. In the overall RRMPs process, the outcome is to optimize and balance the resource arrangement between them. That is the number of the planned radio channel for the voice and data. As more value added services, including mobile commerce bloom, the demand of data service rises up. Also, the new generation network 3G will also occupy the usage of GSM network since there is internal roaming from 3G to 2G network, or 2G network to 3G network due to coverage issue. Both will contribute to the increase demand on 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 channel not only affects the probability of the customer making a successful call, it also affects the 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. If a cell does not have enough radio resource for the cells to hand in, the handover failure of the original call will occur. The worst case happens when the handover is triggered by coverage or quality problem (coherent channel or adjacent channel interference). If the call cannot handover to neighbor cell instantaneously, the call will have a chance to drop. The designed GOS value will affect the number of radio channels or the radio resource, in turn that will affect the network performance and m-payment application services.

Data mining techniques are used to investigate the relationship between each factor, and especially to study the property of the measured GOS with other key factors which affect the network performance.

Once the data is mined, a number of rules will be set for maintaining the GSM network performance. If the GOS value is under a certain threshold, the other key factors such as the call setup successful rate or the call drop rate will also fall within an acceptable range.

Data are collected in hourly and daily basis at regional level or cell level. Data are generated and distributed from system-wide data to individual cell in order to discover unusual patterns. They will have to discover the cause of irregular pattern by investigating the origin of unusual network behaviors. Cell level data are analyzed to discover unusual pattern affecting the grade of service and radio resource.

4. RESULTS OF MINING GSM NETWORK DATA

Different data mining methods are applied in order to discover interesting patterns at RRMPs, they include visualization, clustering, association rule and association rule for classification. Map visualization is used to maintain base station performance to see if any abnormal event occurs. The following mining result based on historical data can search some useful and interesting patterns, trend, information etc. This knowledge can help in dimensioning the network loading and maintaining the network performance. Castaneda Data Mining Tools [9] are used to find and retrieve interesting patterns from our raw data for further analysis. The findings are very comprehensive. However, some samples are given here to for illustration.

4.1 Map 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., its distribution among the proximity of Macao. Also, we can assign other attributes for displaying the result on a geographical map. For traffic (busy_tch_mean), Figure 4 shows the distribution of traffic density in different areas. For the Tch Gos, it shows which area has higher Tch gos, etc.

The visualization results can be useful for traffic dimensioning and resources management, and it can easily alert us on traffic problems and aid in shifting some high traffic area to other areas (from Red to Green) in order to balance the 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, and by alerting this, network operators can design a better network to ensure have a good quality and performance of base stations.

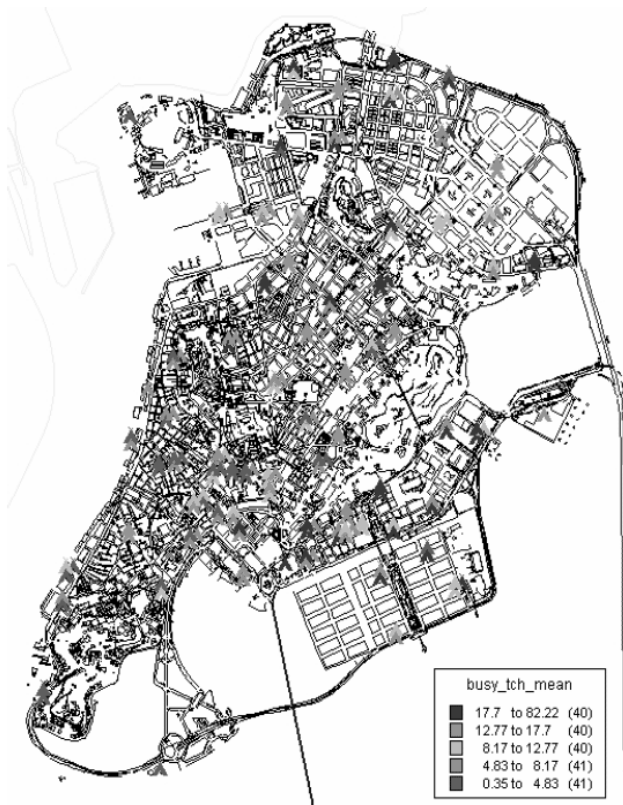


Figure 4 Traffic-Macau

4.2 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. In Figure 5 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 equipments in keys dates such as days of week with high traffic.

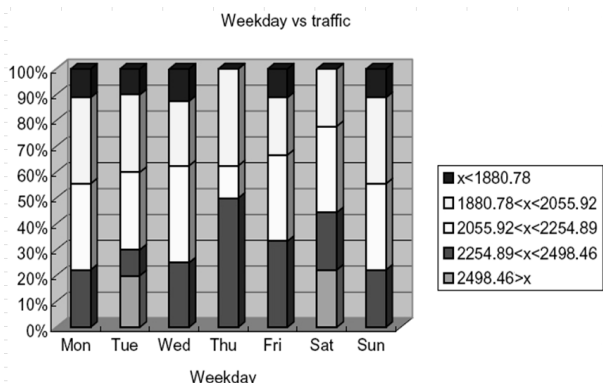


Figure 5 Traffic distributions by weekday

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

4.3 Clustering

Using clustering, it can give us a general idea of the insight of the data when there are too many input attributes. This can help to clarify the healthy status of the system at different levels, such as: system daily, system hourly, and district daily and district hourly level. By analysis the following data, we can have a general idea of the status of the system. For example, in Figure 6 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.

System cluster

Attribute	Values	Cluster 0	Cluster 1	Cluster 2
weekday	Friday	0.176471	0.115385	0.157895
	Monday	0.058824	0.192308	0.157895
	Saturday	0.352941	0.038482	0.105263
	Sunday	0.352941	0.038482	0.105263
	Thursday	0	0.192308	0.157895
	Tuesday	0.058824	0.289231	0.105263
	Wednesday	0	0.153946	0.210526
Csfr_discr	4.29910628019324_x_4.58421739130435	0.411765	0.346154	0.368421
	4.58421739130435_x_4.87733333333333	0.235294	0.038482	0
	4.87733333333333_x_5.14583333333333	0.058824	0	0.105263
	5.14583333333333_x_5.7325	0	0.078923	0
	5.7325_x_5.7325	0.058824	0.038482	0
	5.7325_x_5.7325	0.235294	0.5	0.526316
	x_4.29910628019324	0	0	0
Dcr_discr	1.36118883118883_x_1.43906565656566	0.294118	0.423077	0.315789
	1.43906565656566_x_1.51861111111111	0.529412	0.038482	0.421053
	1.51861111111111_x_1.61	0.117647	0	0.210526
	1.61_x_1.61	0.058824	0	0.052632
	x_1.36118883118883	0	0.538482	0
		0	0	0
		0	0	0

Figure 6 System Cluster

4.4 ASSOCIATION RULES FOR CLASSIFICATION

Do not include headers, footers or page numbers in your submission. These will be added when the publications are assembled.

We try to find use the rules that control the network performance for maintaining 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. 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 $y < x < z$ (Table 1). An important task at RRMPs is traffic profile analysis based on current cell site traffic and by considering the mined results as in Figure 7

5. CONCLUSION

As mobile network will have to support different kind of value added service including commercial mobile payment applications, the resource planning and performance become important. Traditionally the network statistics are monitoring by a system-calculated figure and the irregularities are analyzed manually by a group of technique people. By applying data mining techniques, interesting pattern, rule, or any other kind of presentation of data can be retrieve. The operator or radio planning engineer can obtain valuable information from it.

Table 1 Csfr Rules Classifier

<p>Description</p> <p>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.</p> <p>Usage: Quality of service, Fault finding, Service accessibility; Basis: Cell</p>

IF sdech_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 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 THEN csfr_discr=14.8159585342334_x_31.6589051233397

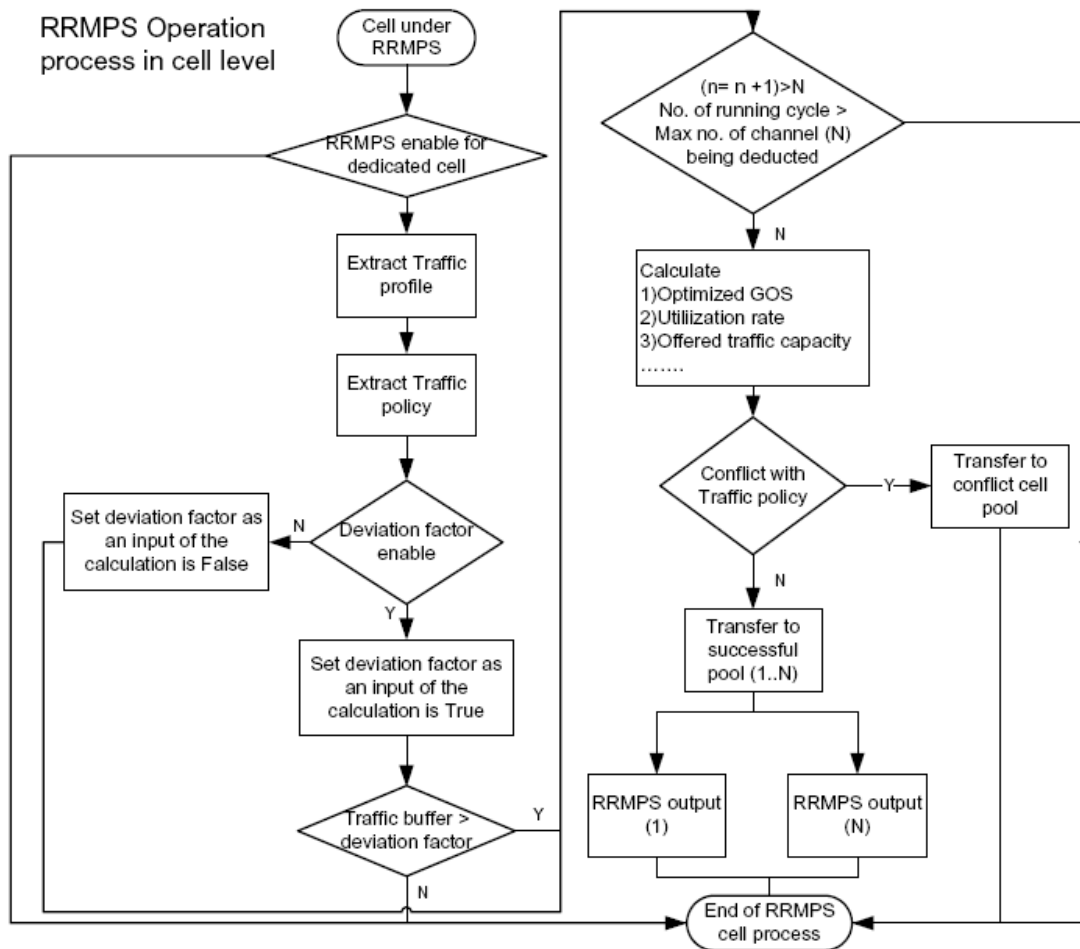


Figure 7 Flow chart cycle for RRMPS calculation

By means of map visualization the traffic channel monitoring is easy 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. Several other attributes are also put into analysis; not only the traffic blocking but also the key indicators such as drop call and call setup failure. Clustering puts the 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.

With the aid of data mining, the network performance could be more effectively monitored, and from which provides a better configuration. This also gives the 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 for both 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 the short message service at OMCS side, i.e. SMS Center Performance, SMS Successful Rate, SMS System Error Rate, SMS Absent Rate, SMS Delivery Success Rate, and SMS Daily failure.

Keep 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 such as the SMS credit payment.

6. ACKNOWLEDGMENTS

The authors are grateful that this work is supported by the Research Council, University of Macau.

7. REFERENCES

- [1] S. Fong, E. Lai, 2005. "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, 9-12 May, 2005, p. 1106-1116
- [2] Pedersen, P.E.; Methlie, L.B.; Thorbjornsen, H, 2002. "Understanding mobile commerce end-user adoption: a triangulation perspective and suggestions for an exploratory service evaluation framework", System Sciences, HICSS. Proceedings of the 35th Annual Hawaii International Conference on 7-10 Jan 2002 Page(s):8 pp.
- [3] Anna Hać, Abhinay Armstrong, "Resource allocation scheme for QoS provisioning in microcellular networks carrying multimedia traffic ", International Journal of Network Management, Volume 11 Issue 5, John Wiley & Sons, Inc.
- [4] Ing-Ray Chen, Naresh Verma, 2003. "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
- [5] Fredrik Berggren, Remco Litjens, 2006. "Performance analysis of access selection and transmit diversity in multi-access networks", MobiCom '06: Proceedings of the 12th annual international conference on Mobile computing and networking, September 2006.
- [6] Jane Y. Yu, Peter H. J. Chong, Ming Yang, 2006. "Performance of Microcell/Macrocell cellular systems with reuse partitioning", Mobility '06: Proceedings of the 3rd international conference on Mobile technology, applications & systems, October 2006
- [7] Smith, P.J.; Sathyendran, A.; Murch, A.R., 1999. "Analysis of traffic distribution in cellular networks", IEEE 49th Vehicular Technology Conference, Volume 3, 16-20 May 1999 Page(s):2075 - 2079
- [8] Walsh, S.; Garcia, E.; Sezer, S., 2005. "Knowledge-based wireless fair queuing using a traffic-profile compensation technique for broadband fixed wireless applications", The 61st IEEE Conference, Vehicular Technology, 2005, Volume 3, 30 May-1 June 2005, Page(s):1913 - 1917
- [9] Castaneda DMS, 2004. <http://www.girgese.com>, 19 Sep