A Dynamic Bandwidth Allocation Scheme for Interactive Multimedia Applications over Cellular Networks

Kirti Keshav
Samsung India Software Operations Pvt Ltd.
Block B, No 66/1 Bagmane Tech Park
C V Raman Nagar, Bangalore, India
kirtik@ece.iisc.ernet.in

Pallapa Venkataram
ECE Department
Indian Institute of Science
Bangalore, India
pallapa@ece.iisc.ernet.in

***Abstract*—Cellular networks played key role in enabling high level of bandwidth for users by employing traditional methods such as guaranteed QoS based on application category at radio access stratum level for various classes of QoSs. Also, the newer multimode phones (e.g., phones that support LTE (Long Term Evolution standard), UMTS, GSM, WIFI all at once) are capable to use multiple access methods simultaneously and can perform seamless handover among various supported technologies to remain connected. With various types of applications (including interactive ones) running on these devices, which in turn have different QoS requirements, this work discusses as how QoS (measured in terms of user level response time, delay, jitter and transmission rate) can be achieved for interactive applications using dynamic bandwidth allocation schemes over cellular networks. In this work, we propose a dynamic bandwidth allocation scheme for interactive multimedia applications with/without background load in the cellular networks. The system has been simulated for many application types running in parallel and it has been observed that if interactive applications are to be provided with decent response time, a periodic overhauling of policy at admission control has to be done by taking into account history, criticality of applications. The results demonstrate that interactive applications can be provided with good service if policy database at admission control is reviewed dynamically.**

Keywords-Cellular networks; interactive applications; dynamic bandwidth allocation.

I. INTRODUCTION

As the bandwidth available to the wireless network users increases, applications (APPs) are becoming more bandwidth hungry. Apart from the bandwidth, other main requirements for current day interactive applications are tight temporal relations, less loss rates and energy efficient data transfer for battery operated hand held devices. This is because session time of interactive applications tends to be longer with intermittent data transfer and in such cases mobile phone's energy consumption should be minimized. Various radio resource management (RRM) techniques has been developed for provisioning QoS for wireless cellular networks. Usually network operators chooses proper access methods and proper RRM policies for multi mode mobile phones.

Interactive traffic classes are meant for bursty, intermittent data transmission. During interactive session, key factor is the response time to users at both the ends. Amount of data transfer could be huge, but it would be bursty in nature. Examples of such applications include instant messenger for voice, chat & video, placeware (conference), netmeeting whiteboard, video messaging, online trading or transactional data client/server applications (such as SAP's, Peoplesoft's and Oracle's), and telnet. Another example of interactive applications is Navigation using off board method, in which the map to be presented on user's screen depends upon, which direction user would take on road. Apart from these applications, many time critical application such as in healthcare sector wherein the doctor needs to instruct interactively for surgery remotely to his colleagues or systems at far off distance. In summary, the main characteristic of interactive class traffic is the strict response time requirement with interactive class traffic while other traffic types are not so delay sensitive.

There are major complexities involved for running real time interactive multimedia applications over cellular networks. Few examples are given here. As the network delays are not predictable, many time critical applications like online trading can suffer and can cause financial losses to users. Similarly for interactive wireless games, if bandwidth and tight temporal requirements are not met, the whole gaming experience can be of very poor quality. For interactive news, if temporal relations between audio/video are not maintained, then it can lead to poor user experience.

Apart from the requirement of necessary required data rate, temporal requirements and necessity of mitigating effects of wireless problems such as fading, frequent handovers due to mobility pose the major challenge for seamless interactive multimedia experience. With the increasing demand for multimedia services in wireless networks, a great deal of attention needs to be paid to resource allocation for time critical interactive multimedia application data and on ways to provide seamless multimedia access in the next generation mobile communication networks. Problem of intelligent usage and allocation of the available bandwidth

 ISBN:978-1-61208-113-7

in a wireless environment is still a challenge due to client mobility and radio born errors.

In the remaining of this sections, QoS requirements of interactive applications and problems faced by interactive applications in cellular network environment are discussed. Following the discussion about related work in Section II, in Section III, the proposed bandwidth allocation scheme is pre- sented by giving details of components present in base sta-‘ tion. After providing detailed algorithm for selecting pri- orities according to host of factors to provide for reliable service to interactive applications, in Section IV, the simulation environment is presented. Thereafter, details of simulation procedure employed in designing of discrete event simulator are presented in Section V. Modelling aspects for various technologies such as LTE (Long Term Evolution), WCDMA (Wideband Code Division Multiple Access) are presented in Section VI. Sections VII and VIII contain details of results and conclusion, and finally, in Section IX, direction for future works is provided.

A. QoS requirements for interactive applications

Interactive class applications mainly comprise of a human or a machine or a remote server at both the ends of communication link and having a response time constrained data transfer. It is characterized by the request response pattern of the end user. Also interactive applications are generally foreground operations in which user waits for the operation to complete before proceeding further and have very tight response time (latency) requirements. Multimedia data can include text, graphics, images, audio and video in various combinations. By strict temporal relations it is meant that the different data types must be received and presented to user as per critical timing relationships. An entity at the destination is usually expecting a response message within a certain period of time. Therefore, the round trip propagation delay (RTD) and delay jitter are the key requirements for such applications [9]. The examples of typical interactive applications such as interactive games and interactive class room illustrates such requirements. For example, interactive games are the one which use the network to interact with other users or systems. Although requirements would depend upon specific application, bandwidth, delay, delay jitter, response time are the key parameters. Many interactive applications try to exchange high volumes of data, but demand very short delays and response time of 100-250 ms is a typical requirement.

B. Problems in using interactive applications over cellular networks

The data encoding methods affect the experience of interactive applications and proper encoding methods should be choosen carefully. Various advances has occurred in data representation schemes for multimedia (MPEG-2, MPEG-4 etc). MPEG-1 and MPEG-2 employ frame based coding techniques in which each rectangular video frame is treated as a unit for compression. The main concern was high compression ratio and satisfactory quality of video under such compression techniques. MPEG-2 had small interaction and is therefore not much helpful for interactive features of communication over cellular communication. MPEG-4, besides compression enables features required for user interactions. It adopts to a new object based coding approach - media objects are now entities for MPEG-4 coding. Media objects can be either natural or synthetic. The bandwidth requirement vary from 5 Kbps to 10 Mbps, so MPEG-4 is highly suitable for interactive multimedia.

To support various kinds of quality of service (QoS) requirements for interactive multimedia in wireless networks, resource provisioning is a major issue. Call admission control (CAC) and dynamic bandwidth management are such provisioning strategies to limit the number of call connections into the networks and to down grade other low priority classes in order to reduce the network congestion and call dropping.

In wireless networks, another dimension is added: call drop due to high users mobility. A good CAC scheme has to balance the call blocking and call dropping in order to provide the desired QoS requirements. Reservation of resources in the network is required to help smooth working of interactive multimedia. This also requires coordination amongst all parties such as admission control, policing, etc.

II. EXISTING RELATED WORKS

One can find lots of work in the literature and standards which talk about providing QoS to various classes of traffic in wireless networks.

One of the most popular strategies for wireless mobile multimedia networks which serve different types of customers with differing bandwidth requirements is the reserve channels (RC) CAC strategy. Call admission control schemes can be categorized based on their handoff-priority policy or queuing priority schemes [5] [6].

1. Guard channel (GC) schemes: In this type of scheme, some channels are reserved for handoff calls. Four different types of CAC schemes have appeared in the literature: Cutoff priority scheme, fractional GC scheme, rigid division-based scheme and new call bounding scheme.

2. Queuing priority (QP) schemes: In this type of scheme, calls are accepted whenever there are free channels. Depending on the approach, new calls are blocked and handoff calls are queued, vice versa, or all calls are queued and the queue is rearranged based on certain priorities.

The work presented in this paper falls conceptually within the second category, that is, of the QP schemes, as in our CAC scheme, both handoff calls and the new calls originating from within the cell are accepted if enough free channel bandwidth exists to accommodate them, and no portion of the bandwidth is restricted for access of either

 ISBN:978-1-61208-113-7

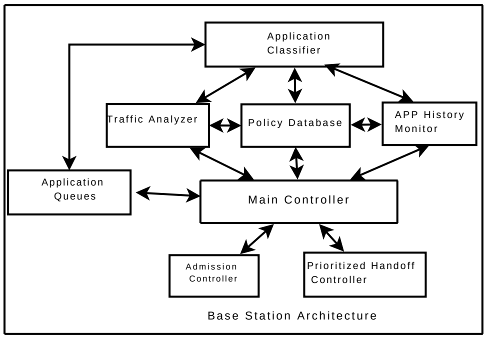

Figure 1. Architecture of bandwidth allocation scheme at base station

type of call (therefore, our scheme is conceptually similar to [8]).

III. Proposed bandwidth allocation schemes

In this paper, we propose a scheme to be considered at network side, to effectively allocate and maintain bandwidth in such a way that QoS requirements of various applications do not suffer and at the same time, deserving interactive applications get their bandwidth allotment easily and the network maintains the interactiveness with high fidelity.

A. Architecture of the bandwidth selection scheme at a base station

The architecture to be used by the base station (or Node B, e-NodeB, whichever is applicable based on communication system) to implement proposed bandwidth selection scheme is depicted in Figure 1 and as can be observed the main components are Application History Monitor, Traffic Analyzer, Application Classifier, Policy Database, Admission Controller, Prioritized Handoff Controller and Application Queues.

In the proposed system, base station maintains the QoS for interactive applications using admission control and traffic shaping operations based on dynamic policies which are derived from history of traffic using traffic analyzer and application history monitor component. Based on regular updating of policy database, intelligent maintenance of application queues, admission control and prioritized handoff control operations, an optimal traffic level is maintained in the system.

Here is the brief description of these components.

Application Classifier: Whenever a user invokes an application, application would request a particular QoS for channel access. Based on QoS parameters of these APPs, and other related parameters applications are classified at two levels - group category level and intra-group category level. Group category level is the broad classification while in intra group category, applications are categorized further inside groups and are assigned priorities.

Application Queues: This component stores the the various arriving applications and considers restrictions such has whether application QoS allows for waiting and channel capacity.

Traffic Analyzer: This component monitors the traffic over a long period of time and maintains the statistics such as application wise drop rate, degradation of QoS, etc. This component along with application history monitor, helps in predicting the application behaviors in better way and hence helps in adaptation of BSC according to current traffic profile.

Policy Database: This component has a database of dynamic policies, which helps in working of admission controller and handoff controller. Policies are dynamic in the sense that a more optimized set of policies are used dynamically based on result of monitoring of traffic by the traffic analyzer and application history due to previous set of policies over a period of time. This means that effect due to one set of policies is analyzed and based on feedback, the policies are updated dynamically for new arriving applications.

Main Controller: This is the central controller in the base station which maintains the central thread in the system. It negotiates with the core network component such as RNC, MSCs and other BSCs. It coordinates application queues, policy database, admission controller and handover controller.

Admission Controller: This component decides the fate of applications when they arrive. If channel doesn't have sufficient suitable amount of bandwidth, when application arrives, a decision to block or drop the call is taken here. This component works in tandem with policies component and main controller. Specially for interactive applications, more restricted admission control mechanism is used. Even if manageable bandwidth is available, an interactive application is admitted only after making sure the current application would survive the predicted scarcity of bandwidth in near future based on experience from previous application history.

Handover Controller: This component takes a decision regarding handover of current application to another cell, if current cell is totally occupied and neighbor cell can provide the required QoS. It also checks the direction of movement (to be negotiated through the main controller from other BSCs and RNC), serving cell signal strength, neighbor cell capabilities, application history for making the appropriate handover commands to users. This feature is very helpful especially in off-board Navigation, wherein application has to pre-download the map from server, based on the prediction of route which user might adopt.

Application History Monitor: This component monitors at detailed level, the history of application calls at the local level (BSC).

In this scheme priorities are assigned in a very dynamic way. Proposed scheme involves monitoring of various call performance parameters in every fixed interval and after

 ISBN:978-1-61208-113-7

Algorithm 1 Algorithm of proposed scheme

Begin
Input: Application is to be dequeued from the APP wait List for admitting in system
Output: APP to admitted - Yes/No ?
Initial priorities assigned to various APP categories based on APP characteristics
At every departure, priorities are reviewed
APP with maximum priority are chosen
if New APP doesn't meet bandwidth Requirement **then**
    Do not drop the APP, wait for next departure. max wait time depend on APP.
else
    Admit the APP
end if
Next admission decision taken at next departure
End

Algorithm 2 Simulation - Priority review algorithm

Begin
Input: Periodic review timer has expired
Output: New updated policies
Check and consult traffic analyzer, policy database, APP history monitor, interactive application response time.
Update priorities to be used in APP Wait Queue accordingly.
End

observing key characteristics a fresh decision regarding priorities for the next interval is taken. These priorities are then taken into account for decisions such as admission, dropping, reducing bandwidth of already accommodated application. Example of various decision factors which can be used in deciding priority are:

a) History of traffic pattern during the day: Traffic pattern in a given time interval of the day.

b) Geographical Area: (i) e.g., if it is hospital - voice traffic is generally given highest priority, however interactive health care applications are also taken care with more priority. (ii) for residential area, applications such as interactive games and shared video news are given more priority.

c) For office areas, application flows involving stocks, email, video conferencing are given high priority.

d) In the rural areas, interactive class room and other applications basted on history are decided to be of higher priority.

e) While in UMTS case, priorities are changed dynamically based on interference, in case of LTE, priorities of applications are updated dynamically based on location in the cell (cell edge users etc). For GSM cells, for various application types, differentiation is not done within cell.

Thus in proposed scheme, always a fare treatment to applications is done by network. Our scheme is shown, via an extensive simulation study comparison and a conceptual comparison, to clearly excel in terms of QoS provisioning to users for interactive applications while balancing overall network parameters such as overall reduction in cellular interference, overall throughput, edge-user throughput, cell user blocking probability and cell capacity . To the best of our knowledge, this is the first work in the relevant literature where such an approach has been proposed.

IV. Simulation Environment & Procedure

For simulation, discrete event simulation methodology has been employed. In the simulation, application arrivals are assumed to be random, and are considered as poisson arrivals. When an arrival occurs, it is classified into one of the application types as mentioned in the table 1 and whether it is same cell generated or its a handover-ed application from another neighboring cell is decided.

Application category is decided randomly using a uniform random variable. Different application types are assigned higher/lower rate according to the arrival rate desired for simulation. For example if APP1 type of application are desired to be of higher arrival rate type, then they are assigned more range in the uniform random variable output while determining their probability of arrival.

Algorithm 1 describes the procedure of admitting new application in the system. Based on application's type, characteristics such as bandwidth required and time it will stay in the cell are calculated. In the simulation, while bandwidth required are hard coded, stay time in cell is estimated using exponential distribution, with each application type having different mean stay time. During simulation, interactive applications are assigned longer call duration as shown in Table 1.

After determining application characteristics, available bandwidth in the channel is checked by the cell's bandwidth allocation module. If this module detects that it can accommodate the currently arrived call without affecting its current application performance, then the call is admitted. If however, bandwidth allocation module finds that the arriving call cannot be admitted, then the call is put into queue. As shown in Algorithm 1, applications are not immediately dropped if bandwidth is not sufficient at the time of new application arrival, rather a fresh assessment of situation is done at every departure of application to know if the leaving application has freed the necessary bandwidth in which new application can be accommodated. Each application category has separate queue having different sizes, with sizes defined by mean bandwidth requirement of respective application category. Each of this queue is assigned with a priority, which depends on many factors such as response time, application's history and criticality of application. Each of

Table I
CHARACTERSTICS OF APPLICATION TYPES

	Reqd APP bandwidth(Kbps)	Mean call duration	Max wait time	Percentage of App	Example
APP1	20	5	1	40	Voice Call
APP2	40	2	2	10	InterAct Video News
APP3	10	1	1	20	InterAct Share Mkt Trading
APP4	60	10	1	30	InterAct Wireless Games

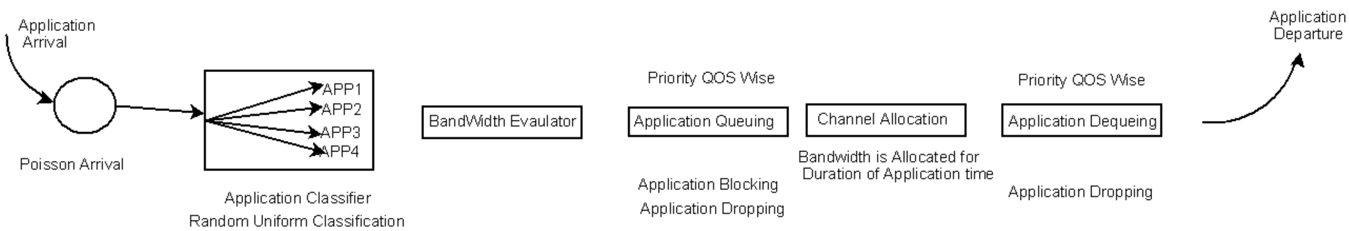

Figure 2. Simulation Model

the application type has a maximum wait period, on whose expiry, if application is still not scheduled to be taken into channel, the application is dropped.

Now whenever, an application leaves the channel, the bandwidth allocator sees if some applications are pending in the queues. It checks the queue priority wise, with higher priority given to more real time interactive application categories and with those APPs having less wait time. Hence by controlling priority of queues and wait time of different queue types, QoS for interactive applications can be guaranteed. Applications flow characteristics are depicted in Table 1. As described above, simulation environment is simulated by performing simulations as per the simulation model depicted in Figure 2. Special care is taken for interactive applications. In today's smart phone, various types of applications are simultaneously running. However it is the interactive applications, which demands most real time behavior from network. For this reason, as described in Algorithm 2, policies are revised periodically based on consultation from traffic analyzer, APP history monitor and interactive application specific response time. Accordingly whenever the network sees that new application type is of interactive type, the main controller in our proposed scheme, learns from previous experience by discussing with App history monitor to check if the available bandwidth with Network is sufficient for future needs of this type of Interactive application. So in essence, apart from assigning higher priorities to interactive applications, our scheme takes care of smooth running of interactive applications by learning in advance about the future need of similar application. This learning of behavior of various applications is done for all users of Network and a central database about application history is maintained. Every e-NodeB/RNC consults this database and takes action at NodeB level at the time of admission of new interactive applications.

V. MODELING ASPECTS

The actual radio resource management (RRM) algorithm definitely depends upon the technology been used at physical layer. For example, if a CDMA- or WCDMA-based (3G) system is considered, then, the fact that system capacity of radio channels (e.g., in 2G systems), but is rather affected by so called "Soft Limited" interference based CDMA behavior has to be kept as a criteria for RRM algorithms. Similarly, if OFDMA is considered at physical Layer (4G, LTE), then apart from usual parameters such as users requirements on the radio and the channel condition, the location of the user in the target cell becomes a significant factor. This is because in LTE, because of OFDMA, there is almost nil intra cell interference, but a high inter-cell interference exist as cell edge users share the same bandwidth in neighboring cell. So to provide good QoS to interactive applications in terms of upper limit on response time, application specific needs have to be taken care by bandwidth allocation and traffic shaping schemes in a cross layer way.

In WCDMA case, the parameters the energy per data bit to effective noise power density ratio $\frac{E_b}{N_t}$ and data rate R_b help in decoupling load control with physical layer parameters [3], [4]. It turns out that relationship between data rate and resource consumption is non linear. This is because resource consumption also depends on QoS requirements [4].

Uplink direction:

$$\left(\frac{E_b}{N_t}\right)_j = \frac{W/R_{bj}.\hat{P}_j}{(1-k).\sum_{i\neq j}\hat{P}_i + I_{other} + I_{th}} \quad (1)$$

W - The chip rate of the system.In UMTS the rate is @=3.84 MChips/S

$\hat{P}_i, \hat{P}_j$- the mean received powers from users i, j respectively.

k - Describes the influence of interference reduction schemes, such as multi user detection (MUD)

 ISBN:978-1-61208-113-7

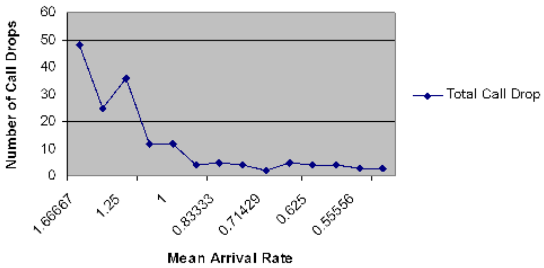

Figure 3. Call Drop Performance in this Scheme

$Iown - \sum_{i \neq j} \hat{P}i$ - the interference from other users in the same cell(cell of user j)

I_{other} - The interference from all users in other cells(other than user j^{th} cell) where the signal is not destined to the referenced Node B (Base station of user j).

I_{th} - The thermal background noise.

The current uplink load can be estimated using $\eta_{ul} = \frac{I_o - I_{th}}{I_o}$, where I_o is the total power received at Base Station.

Downlink Direction:

In the downlink direction, similar analogous results can be derived with the difference that in downlink, one Node B transmits data to many mobile users (mobile stations) and reference point for the downlink investigations is at the user j. The current downlink load can be estimated using $\eta_{dl} = \frac{I_o - P_p}{I_o}$, where I_o is the total power received at base station.

Similarly for LTE, the model followed is similar to "Softer Frequency Reuse based Resource Scheduling Algorithm", as mentioned in [7], in which cell edge users are having greater probability to use the frequency band with higher power and the cell center users are having the higher probability of using frequency band with lower power.

VI. Results and Discussion

Using the above mentioned method of periodic overhauling of policy for admission and traffic shaping for dynamic bandwidth management, after simulations it is seen that apart from being able to provide better QoS, response time to interactive applications, the over all call drop is also kept in check (Figure 3). Since the proposed scheme involves monitoring of various call/application performance parameters every fixed interval and since after observing key characteristics a fresh decision regarding priorities for the next interval is taken, the result obtained suggest that for interactive applications, dynamically adapting of policies is must in order to give good experience to interactive application users while maintaining the QoS of background services under check.

VII. Conclusion

As we have seen in above sections, the proposed bandwidth allocation scheme is an efficient one and when compared to other works, its performance is proved to be better. This shows that in order to provide time critical response to interactive users in reliable way, bandwidth management policies should be reviewed periodically and should take into account the response time, traffic pattern, previous call drop rates, application history (which applications are critical and which flows have given maximum profit in past to network operator).

VIII. Future Work

We are working on design of comprehensive framework to include all aspects of radio resource management such as admission control, scheduling, policing, etc., so as to provide best service for interactive applications.

References

[1] Y. Xu, H. Liu, and Q. Zeng, "Resource management and QoS control in multiple traffic wireless and mobile Internet systems", 14th International Conference on Computer Communications and Networks, vol. 17, no. 19, pp. 125-130, Oct. 2005.

[2] A. Kind, X. Dimitropoulos, S. Denazis, and B. Claise, "Advanced Network Monitoring Brings Life to the Awareness Plane" IEEE Communications Magazine, vol. 46, no. 10, pp. 140-146, October 2008.

[3] 3GPP Specifications, Radio Resource Control (RRC) Protocol specification, 25.331, version 3.21.0, Release 99

[4] U. Bernhard, E. Jugl, J. Mueckenheim, H. Pampel, and M. Soellner, "Intelligent Management of Radio Resources in UMTS Access Networks" Bell Labs Technical Journal, vol. 7, no. 3, pp. 109-126, 2003.

[5] Y. Fang and Y. Zhang, "Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks" IEEE Transactions on Vehicular Technology, vol. 51, no. 2, pp. 371-382, 2002.

[6] H. A. Mohamed, "Call Admission Control in Wireless Networks: A Comprehensive Survey" IEEE Communications Surveys & Tutorials, vol.7, no.1, pp. 49-68, 2005

[7] S. Hussain, "Dynamic Radio Resource Management in 3GPP LTE", http://www.bth.se/fou/cuppsats.nsf/all/c858bf5b4979a6b3c1257552004262f6$file/Sajid_Hussain_MSc_Thesis_Report.pdf, last accessed on 1-11-2010

[8] C. J. Haung, W. K. Laib, and Y. L. Yanb, "A self-adaptive bandwidth reservation scheme for sectored cellular communications" Information Sciences, vol. 166, no. 1-4, pp. 127-146, 2004.

[9] B. Zheng and M. Atiquzzaman, "System Design and Network Requirements for Interactive Multimedia" IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 1, pp. 145-153, 2005