# Application Based Utility Adaptation with Prioritized Weight Assignment Strategy for Minimizing Delay in Multimedia Applications

M.Bhuvaneswari

Assistant Professor
Department of Electronics and Communication Engineering
Anna University of Technology,Tiruchirappalli
Tiruchirappalli-620024
Tamilnadu,India

P.Seethalakshmi

Director
Centre for Academic Excellence
Anna University of Technology,Tiruchirappalli
Tiruchirappalli-620024
Tamilnadu,India

***Abstract*— The aim of this paper is to propose a joint cross-layer approach between application layer and MAC layer for enhancing Quality of Service (QoS) for multimedia applications. Bandwidth adaptation scheme based on application utility function is used in the application layer as it reflects the satisfaction degree of end-users with the allocated bandwidth. Prioritized Weight Assignment (PWA) in MAC layer assigns higher weight to prioritized multimedia data based on the allocated bandwidth. The application utility function is the cross-layer parameter which jointly adapts the application layer function in coordination with the MAC layer to minimize the delay experienced by the multimedia data. Simulation results have clearly demonstrated the superior performance of the proposed joint cross-layer schemes such as bandwidth adaptation with Prioritized Weight Assignment (PWA) by comparing them with some existing ones**.

***Keywords- Bandwidth adaptation; Utility function; Prioritized Weight Assignmen; Multimedia applications***.

## I. Introduction

The rapid development of wireless communication technologies and its services has greatly enriched the diversity of wireless applications evolving from the traditional voice service to a wide range of multimedia services including data, voice, and video [1]. Different multimedia services over networks have different bandwidth requirements. For example, applications like audio phone and video conference require strict end-to-end performance guarantees; hence it is crucial for the networks to provide reliable and timely packet transmission. On the other hand, applications such as E-mail and file transfer can adapt their bandwidth to various network loads since they can tolerate certain delays. As a result, providing QoS to multimedia applications according to their bandwidth requirements is becoming an important resource management issue for wireless networks.

Bandwidth adaptation is one of the most promising resource management methods to provide QoS guarantees to multimedia traffic in wireless networks. The main feature of bandwidth adaptation is that it can explore the adaptive nature of multimedia applications and dynamically adjust their allocated bandwidth to deal with network resource fluctuations. Examples of such multimedia services include the International Organization for Standardization’s (ISO’s) Motion Picture Experts Group 4 (MPEG-4) [ISO00-1] and the International Telecommunication Union’s (ITU’s) H.263 [5-6].

### *A. Utility-based Multimedia Traffic Model*

In multimedia wireless networks, different applications have different bandwidth requirements. To provide QoS support to multimedia applications according to their bandwidth needs under the wireless environment featuring limited and varying bandwidth resource, the explicit traffic model is needed to reflect the QoS sensitivity of the applications to bandwidth allocation.So, a utility-based multi-class traffic model is proposed to differentiate multimedia traffic according to their adaptive characteristics.

### *B. Utility Function*

Utility function is defined as a curve mapping the amount of bandwidth received by the application to the performance as perceived by the end-user. The key advantage of utility function is that it can inherently reflect the QoS requirements of the end-to-end user and quantify the adaptability of the application. The shape of the utility function varies according to the adaptive characteristics of the application. According to the bandwidth requirements, multimedia traffic can be classified into two broad classes:

- Class I – real-time traffic
- Class II – non-real-time traffic.

Class I traffic can be further classified into two subclasses – adaptive real-time traffic and hard real-time traffic.

#### 1.*Adaptive Real-Time Traffic*

Adaptive real-time traffic refers to the applications that have flexible bandwidth requirements. Typical examples are interactive multimedia services and video on demand [7]. The utility function of adaptive real-time traffic is modeled as follows:

$$u(b) = 1 - e^{-\frac{k_1 b^2}{k_2 + b}} \qquad (1)$$

where k1 and k2 are two positive parameters which determine the shape of the utility function and ensure that when the maximum bandwidth requirement bmax is received, the achieved utility u max is approximately equal to 1.

#### 2. *Hard Real-Time Traffic*

Hard real-time traffic refers to the applications with stringent bandwidth requirements. Examples include audio/video phone, video conference and telemedicine [7-8]. The following utility function is used to model hard real-time traffic:

$$u(b) = \begin{cases} 1, & when\ b \geq b^{\min} \\ 0 & when\ b < b^{\min} \end{cases} \qquad (2)$$

#### 3. *Non-Real-Time Traffic*

Non-real-time traffic refers to the applications which are rather tolerant of delays. Most traditional data applications such as E-mail, file transfer and remote login [9] [10] belong to non-real-time traffic and they can work without guarantees of timely packet delivery. The following utility function is used to model non-real-time traffic:

$$u(b) = 1 - e^{-\frac{kb}{b^{Max}}} \qquad (3)$$

where k is a positive parameter which determines the shape of the utility function and ensures that when the maximum bandwidth requirement b max is allocated, the achieved utility u max is approximately equal to 1.

### *C. Prioritized Weight Assignment (PWA):*

The data packets that are being transmitted are classified into two major categories as real time and non real time data. Separating the real time and non real time data coming from the source, and effectively sending the real time data alone will reduce the delay of the transmitted real time data in the network [17].

#### 1.*Priority Assignment*

By assigning higher priority to the real time data and queuing those data separately and transmitting the real time data before the non real time data will reduce the loss of the real time data during the transmission and also decrease the delay in the transmission of the data. For separating the real time data the timing sequence of the packet coming from the network layer is considered. Higher weights are assigned to high priority real time data and the other data are assigned lower priority. Thus, the data in the higher priority are sent before the lower priority and thus the loss of data can be reduced up and delay in transmitting the data is reduced to an certain level.

### *D. Bandwidth Assignment*

Thus, for this approach bandwidth to the certain level of the data is assigned as

$$B_i = (P_i / \sum_j P_j).B \qquad (4)$$

where $B_i$ is the bandwidth of the $i_{th}$ data, $P_i$ is the priority of $i_{th}$ data, $P_j$ is the priority assigned to overall data, B is the overall bandwidth assigned.

#### 1.*Weights Assignment*

Like priority, weights are assigned to the packets such that the weighted packets are transmitted first. Based on the priority, more weights are given to higher priority packets and are transmitted first.

While assigning weights, the weights of the higher priority must be higher than the lower priority packets as

$$W_i \geq W_{i+1} \text{ where } i = 1, 2,.. \qquad (5)$$

The assigned weights must be updated such that the data can be transmitted without delay and effective output can be obtained. The updating of the weights can be done as

$$W_1^{(new)} = W_1.(RT_1)^{-\alpha} \qquad (6)$$

$$W_i^{(new)} = \min\left[\left({}^{W_i}\!/_{W_{i-1}}\right)W_{i-1}^{(new)}, W_i.(RT_i)^{-\alpha}\right] \qquad (7)$$

Using the above approach, bandwidth can be assigned to specified data and will allow the data to be transmitted with a collision free medium and thus the delay can be reduced up. The paper is organized as follows with System model in section II, Simulation modeling in section III ,results and discussion in section IV and finally conclusion in section V.

## II. SYSTEM MODEL

In wireless networks with multimedia traffic, each call is assigned a utility function with shape depending on its traffic class. When a call requests a connection to the network, it is assumed to provide the following information [1][9]:

- Traffic class
- Bandwidth requirements
- Utility function.

With adaptive bandwidth allocation paradigm, if there is enough bandwidth available in the network, the call is allocated its maximum bandwidth b max; otherwise, depending on how much the network is overloaded, the call is allocated a bandwidth ranging from its minimum bandwidth $b_{min}$ to its maximum bandwidth $b_{max}$. Although there is still no standard over all architecture for the end-to-end implementation, some partial solutions have been proposed. As introduced in [11] [12] and [13], the adaptation of multimedia applications can be achieved at different Open System Interconnection in wireless networks.

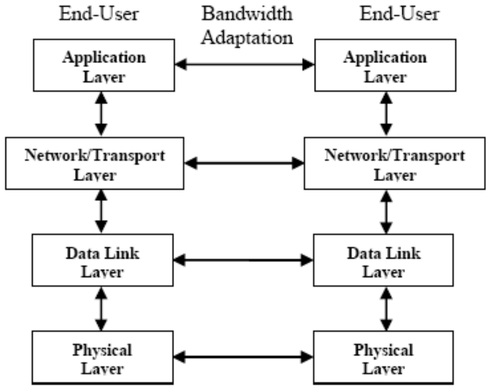


Figure 1: Multimedia adaptation at different OSI layers [12]

The function of each OSI layer is as follows:

- Physical layer**.** At the physical layer, adaptability can be achieved by choosing appropriate modulation techniques e.g. PSK.

- Data link layer**.** At the data link layer, error control mechanisms e.g. retransmission can be used to protect against the varying error rates of wireless links.

- Network/Transport layer**.** At the network/transport layer, routing methods can be used to adapt the applications when there is user mobility.

- Application layer. At the application layer, most multimedia applications can adapt to the changing networking conditions using various multimedia coding techniques [14].

To support the QoS provisioning in wireless networks, bandwidth adaptation often needs to make tradeoffs between multiple QoS objectives. Each call in the network is assigned a utility function to reflect the relationship between bandwidth allocation and the end-user's satisfaction. Bandwidth adaptation is performed in each BS and it consists of two processes –bandwidth degrades and bandwidth upgrades, which are triggered by call arrival events and call departure events, respectively (see Figure 2)

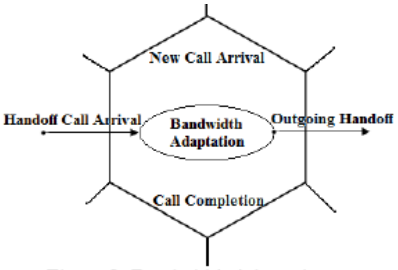


Figure 2: Bandwidth Adaptation

Call arrival events include new call arrival events (a new call is generated within the cell) and handoff call arrival events (a handoff call arrives to the cell). Call departure events include call completion events (a call within the cell completes) and outgoing handoff events (a call leaves its current cell) [15] [16].

### *1.Bandwidth Degrade*

When a new or handoff call arrives to a cell of the network, if the cell has enough bandwidth available, the new or handoff call is admitted at its maximum bandwidth requirement. If the cell is overloaded, the bandwidth of adaptive ongoing calls can be degraded to smaller values to accommodate the new or handoff call.

### *2.Bandwidth Upgrade*

When an ongoing call is terminated due to its completion or outgoing handoff from its current cell to another, if all calls in the current cell have received their maximum bandwidth, the released bandwidth is saved for future use. Otherwise, the released bandwidth can be utilized to upgrade the adaptive ongoing calls that have not received their maximum bandwidth. With the proposed scheme, each call in the network is assigned a utility function.

### *A. Cross-Layer Scheduling Architecture*

The scheduling algorithm at the MAC layer is modeled as an optimization problem with respect to some physical layer constraints and application QoS constraints [1] [17]. At every timeslot, the scheduling algorithm has to produce rate allocation and power for all the k users, which is based on the observation of the current channel state information (CSI) from the physical layer and the queue state information (QSI) from the application layer. Rate allocation and power allocation are selected so that they optimize the system objectives.

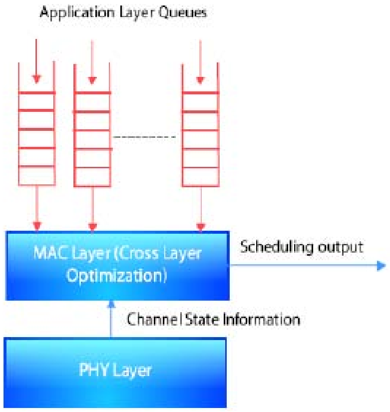


Figure 3: Schematic of a cross-layer scheduling architecture

### *B. Uplink Scheduler Architecture*

The uplink (UL) scheduler [17] comprises three modules: information module, database module and service assignment.

- Information Module

This module performs the following functions and passes them to the scheduling data base module.

- Categorizing the packets on the MS (Mobile Station)basis
- Extracting the queue size information (number of waiting packets and size of each packet of each connection from the BW ( (Bandwidth)-Request messages)
- Deciding the arrival time

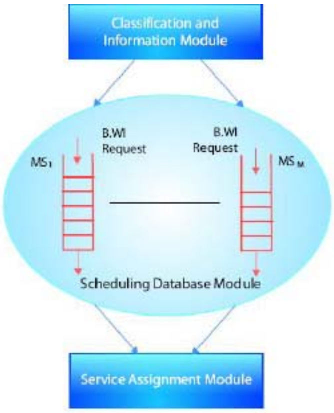


Figure 4: Uplink Scheduler Architecture

The process determines the arrival time of packets that arrived during the previous frame i.e. decides time bound. The time bound is given by the sum of the packet's arrival time and the packet's maximum delay requirement (as determined by the connection QoS Parameters).The type field in the bandwidth request header indicates whether the request is incremental or aggregate. Since piggybacked Bandwidth Requests do not have a type field, PBR shall always be incremental. When the BS (Base Station) receives an incremental bandwidth request, it shall add the quantity of bandwidth request to its current perception of bandwidth needed for the connection. When the BS receives an aggregate bandwidth request it shall substitute its perception of bandwidth needs of the connection with the quantity of bandwidth requested.

- Scheduling Database Module

The scheduling database module serves as the information database for all the MS in the network.

- Service Assignment Module

The service assignment module determines the uplink sub-frame allocation in terms of the number of bits per MS. The number of bits will finally be converted to the number of time slots, i.e. the units used in the information elements (IE) of the Up Link Map (UL-MAP) generation. The number of bits per time-slot is determined by the physical layer of the wireless network.

The surplus bandwidth is distributed among all the connections according to their instantaneous bandwidth requirements. The sharing of the surplus bandwidth among different classes follows priority logic, from highest to lowest: UGS, rtPS, ertPS, nrtPS and BE. In another word the method boils down to the following two stages:

- Reserved bandwidth of each priority class is distributed among the admitted connections of that class, according to the desired scheduling policy.

- The excess bandwidth would be allocated to those connections that have not been granted a part or total of their requested bandwidth.

### C. *Downlink Packet Scheduler*

In this structure, the scheduling scheme is based on a principle called Grant Per Type-of-Service (GPTS). The delay performance is differentiated for each queue and respective time slots are assigned with varying priorities and QoS guarantee [17][18]. The priority function for queue i is defined as

$$\mu_i(t) = \frac{r_i(t)}{R_i(t)/C_i(t)M_i(t)}$$

where *M* is the minimum rate requirement, *Ci(t)* is the number of connections of the *i* queue and *Ri(t)* is the ith transmission capacity at time *t*, which is determined by the channel quality.

Each queue corresponds to one QoS requirement class, respectively. The queue having the highest value of *μ(t)* is scheduled. The estimated average *I* throughput of scheduled queue *i* (denoted as *R (t)*) is updated by the following simple exponential smoothing model:

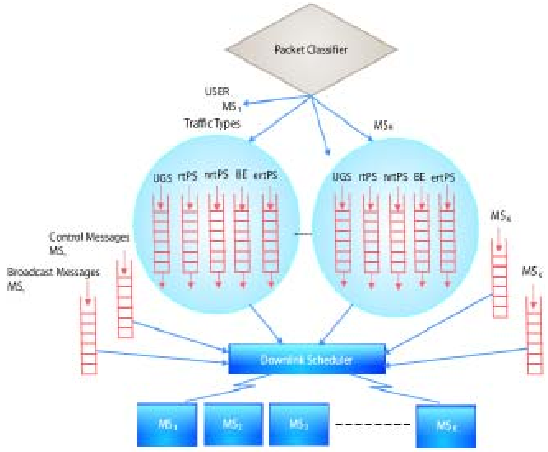


Figure 5: Downlink Packet Scheduler

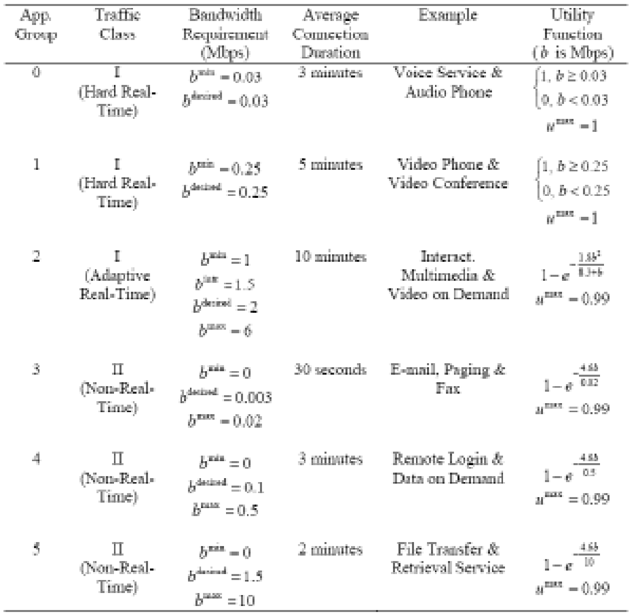
$$R_i(t+1) = (1-\frac{1}{T_i})R_i(t) + \frac{1}{T_i}\sum_{n\in\Omega_k(t)} r_i(t,n) \qquad (8)$$

Where *r (t)* is the supportable data rate for user *i* in sub carrier *n* at time slot *t* and *R(t)* is the average throughput *i i*th received by the i user up to slot *t*. indicate the set of sub-carriers in which user *i* is scheduled for transmission at time slot *t.* In this case, the system throughput is enhanced and efficient bandwidth utilization is achieved.

## III. *SIMULATION MODELING*

To evaluate the performance of the proposed utility-maximization bandwidth adaptation scheme, a multimedia wireless network simulation model has been developed.

**Table 1: Simulator**



| App. Group | Traffic Class | Bandwidth Requirement (Mbps) | Average Connection Duration | Example | Utility Function ($b$ is Mbps) |
|---|---|---|---|---|---|
| 0 | I (Hard Real-Time) | $b^{min} = 0.03$<br>$b^{desired} = 0.03$ | 3 minutes | Voice Service & Audio Phone | $\begin{cases} 1, b \geq 0.03 \\ 0, b < 0.03 \end{cases}$<br>$u^{max} = 1$ |
| 1 | I (Hard Real-Time) | $b^{min} = 0.25$<br>$b^{desired} = 0.25$ | 5 minutes | Video Phone & Video Conference | $\begin{cases} 1, b \geq 0.25 \\ 0, b < 0.25 \end{cases}$<br>$u^{max} = 1$ |
| 2 | I (Adaptive Real-Time) | $b^{min} = 1$<br>$b^{int} = 1.5$<br>$b^{desired} = 2$<br>$b^{max} = 6$ | 10 minutes | Interact. Multimedia & Video on Demand | $1 - e^{-\frac{1.8b^2}{8.3+b}}$<br>$u^{max} = 0.99$ |
| 3 | II (Non-Real-Time) | $b^{min} = 0$<br>$b^{desired} = 0.003$<br>$b^{max} = 0.02$ | 30 seconds | E-mail, Paging & Fax | $1 - e^{-\frac{4.6b}{0.02}}$<br>$u^{max} = 0.99$ |
| 4 | II (Non-Real-Time) | $b^{min} = 0$<br>$b^{desired} = 0.1$<br>$b^{max} = 0.5$ | 3 minutes | Remote Login & Data on Demand | $1 - e^{-\frac{4.6b}{0.5}}$<br>$u^{max} = 0.99$ |
| 5 | II (Non-Real-Time) | $b^{min} = 0$<br>$b^{desired} = 1.5$<br>$b^{max} = 10$ | 2 minutes | File Transfer & Retrieval Service | $1 - e^{-\frac{4.6b}{10}}$<br>$u^{max} = 0.99$ |

## IV. SIMULATION RESULTS

The simulator employs discrete-event simulation to model the traffic and its management in wireless networks.

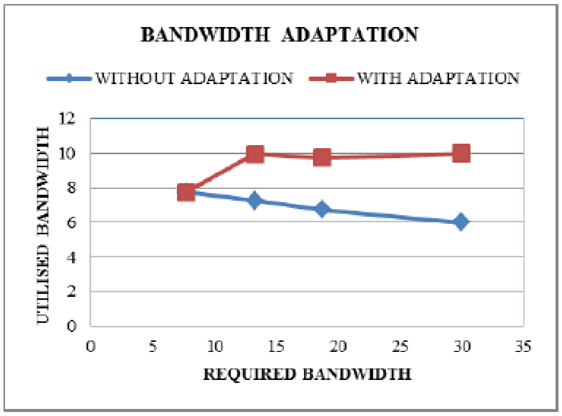


Figure 6: Effect of utilized bandwidth

A discrete-event simulation is one in which the state of the model changes at only a discrete set of simulated time points. With discrete event simulation a real network system is decomposed into a set of separate components. The fundamental element of the simulation is event and the operation of the system is represented as a chronological sequence of events Each event is assigned a time stamp and takes place on a specific component. The result of this event can be a message passed to one or more other components. On arrival at the other components, the content of the message may result in the generation of new events to be processed at some future logical time. The simulation parameters are assumed from [1] and the results are given in Fig.6, 7 and in Table 2 and 3.

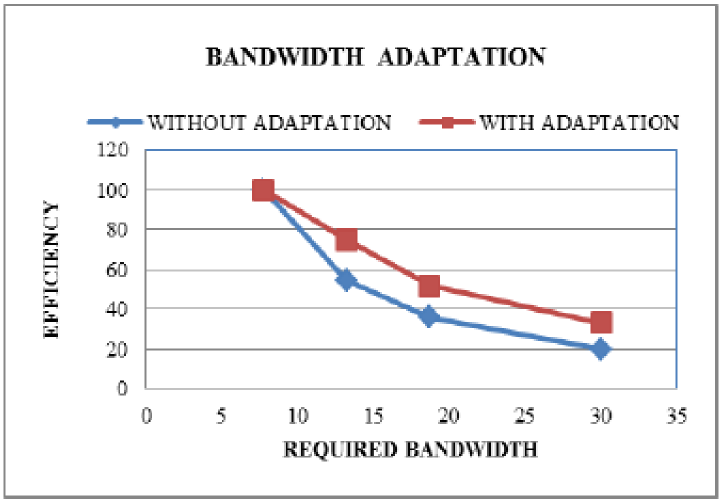
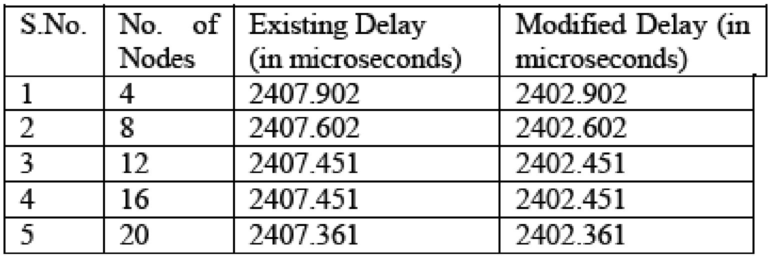


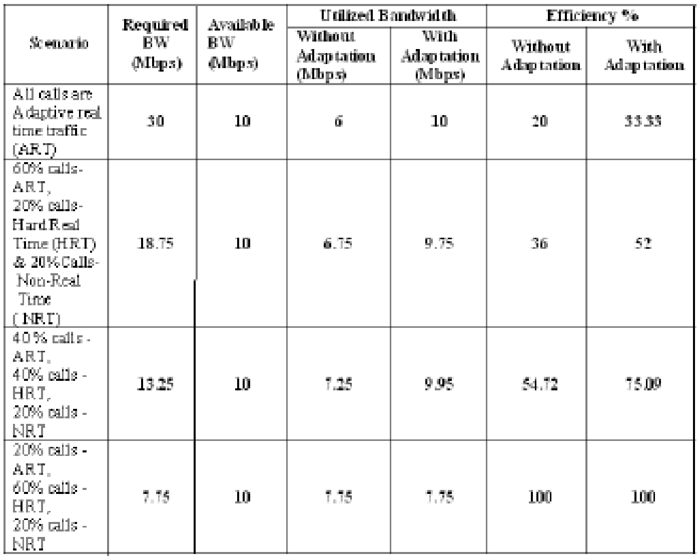
Figure 7: Efficiency of bandwidth adaptation

Table 2 : Simulation Results



| S.No. | No. of Nodes | Existing Delay (in microseconds) | Modified Delay (in microseconds) |
|---|---|---|---|
| 1 | 4 | 2407.902 | 2402.902 |
| 2 | 8 | 2407.602 | 2402.602 |
| 3 | 12 | 2407.451 | 2402.451 |
| 4 | 16 | 2407.451 | 2402.451 |
| 5 | 20 | 2407.361 | 2402.361 |

Table-3: Comparison of delay with the existing model and the modified model



| Scenario | Required BW (Mbps) | Available BW (Mbps) | Utilized Bandwidth | | Efficiency % | |
|---|---|---|---|---|---|---|
| | | | Without Adaptation (Mbps) | With Adaptation (Mbps) | Without Adaptation | With Adaptation |
| All calls are Adaptive real time traffic (ART) | 30 | 10 | 6 | 10 | 20 | 33.33 |
| 60% calls-ART, 20% calls-Hard Real Time (HRT) & 20% Calls-Non-Real Time (NRT) | 18.75 | 10 | 6.75 | 9.75 | 36 | 52 |
| 40 % calls - ART, 40% calls - HRT, 20% calls - NRT | 13.25 | 10 | 7.25 | 9.95 | 54.72 | 75.09 |
| 20% calls - ART, 60% calls - HRT, 20% calls - NRT | 7.75 | 10 | 7.75 | 7.75 | 100 | 100 |

## V. CONCLUSION

Thus, the proposed approach of assigning bandwidth adaptation scheme based on application utility function and PWA to data packets which are transferred from MAC layer efficiently reduces the delay between node to node data transmission. Simulation results have clearly demonstrated the superior performance of the proposed joint cross-layer schemes such as bandwidth adaptation with Prioritized Weight Assignment (PWA) by comparing them with some existing ones.(Fig.8)

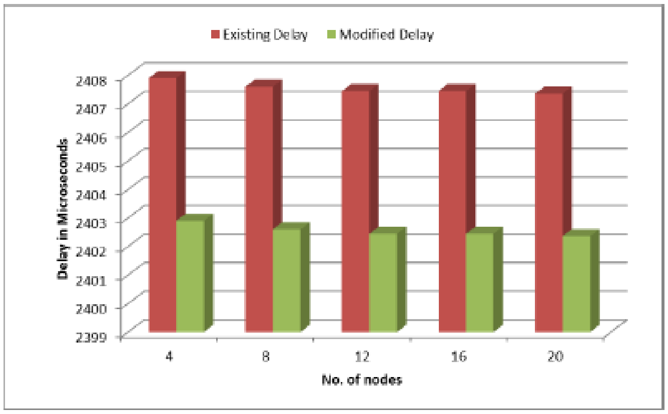


Figure 8: Comparison of delay with the existing model and the modified model

## REFERENCES

[1] G N. Lu and J. Bigham," On Utility-Fair Bandwidth Adaptation for Multi-Class Traffic QoS Provisioning in Wireless Networks," Computer Networks, vol. 51, no. 10, pp. 2554-2564, July 2007.

[2] N. Lu and J. Bigham, "An Integrated Bandwidth Adaptation Schemefor Multimedia Wireless Networks and its Connection-Level Performance Analysis," Journal of Communications Software and Systems (JCOMSS), Special Issue on QoS in Wireless Networks, vol. 2, no. 2, June 2006 (Electronic Publication).

[3] J. Bigham, L. Cuthbert, X. Yang, N. Lu and D. Ryan, "Using Intelligent Agents for Managing Resources in Military Communications," Computer Networks, vol. 46, no. 5, pp. 709-721,Dec. 2004.

[4] A. Alwan, R. Bagrodia, N. Bambos, M. Gerla, L. Kleinrock, J.Short, and J. Villasenor,"Adaptive Mobile Multimedia Networks," IEEE Personal Communications, vol. 3, no. 2, pp. 34-51, April 1996.

[5] V. Rakocevic, J. Griffiths, and G. cope,"Performance Analysis of Bandwidth allocation schemes in Multiservice IP Networks Using Utility Functions," Proceedings of the 17th International Teletraffic Congress(ITC'17), Dec.2001.

[6] N. Banerjee, K.Basu and S.K. Das, "Adaptive Resource Management for Multimedia Applications in wireless networks ," Proc. of the 6th IEEE International Symposium on a world of Wireless Mobile and multimedia Networks(WOWMOM'05) pp.250-257, June 2005.

[7] J.M. Holzman," Asymptotic analysis of proportional fair algorithm", in Proc. IEEE PIMRC 2001, San Diego, CA,pp.33-37

[8] K. Rijkse,"H.263: Video Coding for Low-Bit-Rate Communication," IEEE Communications Magazine, vol. 34, no. 12, pp. 42-45, Dec. 1996.

[9] C. olivera, J.B.Kim, and T. suda, "An Adaptive Bandwith Reservation Scheme for High Speed Multimedia Wireless Networks," IEEE journal on selected Areas in communication, vol. 16, no. 6, pp. 858-874, Aug. 1998.

[10] S. Shenker, "Fundamental Design Issues for the Future Internet,"IEEE Journal on Selected Areas in Communications, vol. 13, no. 7, pp. 1176-1188, Sept. 1995.

[11] J.-Y. Chang and H.-L. Chen, "A Borrowing-Based Call Admission Control Policy for Mobile Multimedia Wireless Networks," IEICE Transactions on Communications, vol. E89-B, no. 10, pp. 2722-2732, Oct. 2006.

[12] A.-L. I. Semeia," Wireless Network Performance Analysis for Adaptive Bandwidth Resource Allocations", PhD Thesis, Stevens Institute of Technology, USA, 2003.

[13] V. Rakocevic," Dynamic Bandwidth Allocation in Multi-Class IP Networks Using Utility Functions," PhD Thesis, Queen Mary, University of London, UK, 2002.

[14] B. Vandalore, W.-C. Feng, R. Jain, and S. Fahmy," A Survey of Application Layer Techniques for Adaptive Streaming of Multimedia," Real-Time Imaging, vol. 7, no. 3, pp.221-235, June2001.

[15] A. Aljadhai and T.F. Znati, "A Bandwidth Adaptation Scheme to Support QoS Requirements of Mobile Users in Wireless Environments," Proceedings of the 9th International Conference on Computer Communications and Networks, pp. 34-39, Oct. 2000.

[16] Y. Xiao, C.L.P. Chen, and Y. Wang, "Optimal Admission Control for Multi-Class of Wireless Adaptive Multimedia Services," IEICE Transactions on Communications, vol. E84-B, no. 4, pp. 795-804,April 2001.

[17] Xiaojing Meng "An Efficient Scheduling for Diverse QoS Requirements in WiMAX", A thesis presented to the University of Waterloo. Waterloo, Ontario, Canada, 2007.

[18] P.Viswanath, D.N.C.Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," IEEE Transactions on Information Theory, vol. 48, no.6, pp. 1277-1294, Jun. 2002

## AUTHORS PROFILE

M.Bhuvaneswari is a Assistant Professor of Electronics and Communication Engineering, Anna University of Technology, Tiruchirappalli, Tamilnadu, India.Her research interests are computer networks, multimedia networks and communication systems..

P.Seethalakshmi is the Director of Centre for Academic Excellence, AnnaUniversity of Technology, Tiruchirappalli, Tamilnadu, India. Her research interests are Distributed computing, computer networks and multimedia networks .