Modified Priority Queuing Packet Scheduling Algorithm for Multi-media Traffic in Wireless Communication Environment

DONG-HOI KIM[1], MYUNG-DONG KIM[1], IL-HWAN KIM[1]

[1]Department of Electrical and Electronic Engineering, College of Information Technology Kangwon National Univ. Chuncheon, Korea.
{donghk, bw001612,ihkim}@kangwon.ac.kr

Abstract: - This paper introduces a new packet scheduling algorithm for multi-media traffic. The proposed modified priority queuing (MPQ) packet scheduling algorithm, realizing the high priority traffic service employs exponential distributions with the higher averages respectively, secures the reduction in the mean waiting time regarding the high priority traffic sensitive to delay. Thus, we could meet the requirements for the mean waiting time of the real and non-real-time traffic services. In the wireless communication system adopting the adaptive modulation and coding (AMC), as the proposed MPQ packet scheduling algorithm assigns the higher order modulation with higher code rates to the higher priority service with higher priority, the algorithm can have done good (work) in increasing the number of multi-media service users satisfying the required quality of service (QoS). The performance study has strongly focused on the effects of the proposed MPQ algorithm.

Key-Words: - Priority Queuing, Adaptive Modulation and Coding, Packet Scheduling Algorithm, Multi-media Traffic.

1 Introduction

Multi-media services with different quality of service (QoS) requirements will be demanded and an algorithm able to increase the number of multi-media service users satisfying the required QoS will be needed. Also, if the algorithm is connected with the physical transmission method such as adaptation modulation and coding (AMC), it can be usefully applied to future wireless systems. For example, in orthogonal frequency division multiplexing Access (OFDMA) environment, time slots or frequency channels as well as power level have become the new resource to allocate. In order to effectively utilize limited resources in the wireless communication environment, AMC have offered an alternative link adaptation method that promises to raise the overall system capacity [1].

AMC provides the flexibility to match the modulation-coding scheme to the average channel conditions for each user. With AMC, the power of the transmitted signal is held constant over a frame interval, and the modulation and coding format is changed to match the current received signal quality or channel conditions. In a system with AMC, users close to the base station (BS) are typically assigned higher order modulation with higher code rates (e.g. 64 QAM with R=3/4 turbo codes), but the modulation-order and/or code rate will decrease as the distance from Node B increases. AMC is most effective when combined with fat-pipe scheduling techniques such as those enabled by the Downlink Shared Channel. On top of the benefits attributed to fat-pipe multiplexing [2], AMC combined with time domain scheduling offers the opportunity to take advantage of short term variations in a UE's fading envelope so that a UE is always being served on a constructive fade. On the other hand, if higher order modulation with higher code rates is assigned to service with higher priority, the number of higher priority service users with the strict QoS requirement will be increased.

In this paper, the proposed modified priority queuing (MPQ) packet scheduling algorithm provides the high priority traffic service with exponential distributions having the higher averages respectively, in order to secure the reduction in the mean waiting time regarding the high priority traffic sensitive to delay. Thus, we could meet the requirements for the mean waiting time of the real and non-real-time traffic services. That is, as multimedia services that require various QoS have been realized, many methods for fairly allocating bandwidths of output links of the packet schedulers for efficiently using the limited radio resources have been proposed. This paper is organized as follows. Section 2 introduces the proposed MPQ packet scheduling algorithm. Section 3 and Section4 provide the numerical analysis and result of the proposed MPQ algorithm for QoS guarantee in multi-media traffic mobile system. Concluding remarks are given in Section 5.

2 Proposed modified priority queuing (MPQ) scheduling

Seeing that the conventionally used priority queuing (PQ) packet scheduling algorithm assigns priority to real-time traffic service first and then carries out the traffic class with the next priority, for the case the real-time packet service is terminated and the buffer has no data [3], but the proposed MPQ packet scheduling algorithm makes difference by applying higher service rate (providing AMC with the higher order modulation with higher code rates) to the traffic class with higher priority than that with lower priority. Fig. 1 illustrates the way MPQ goes. In Fig. 1, when a new packet enters, priority will be set in order according to its class. In case of the real-time traffic, the packet will be accumulated in the high priority queue, but for the non-real-time traffic case, it is done in the low priority queue.

This mechanism allows providing differentiated QoS where high priority traffic will experience less delay and is less likely to be dropped. Clearly, a network offering this service model has to deploy a pricing scheme where user are charged based on the priority of their class; otherwise all the users would submit their traffic in the highest priority class. We consider a pricing scheme where the network announces to users a menu of static prices associated with the different priority classes. Users are free to choose the priority of their traffic, but are billed by network accordingly. In this context, pricing becomes effectively a resource allocation mechanism: one would expect that users assign traffic to priorities in order to meet their QoS requirements at the lowest possible cost, and therefore use high priorities only for traffic with strict QoS requirements. In this study, we investigate this price-based resource allocation scheme called as MPQ and how radio resources are shared among users.

In order to analyze the performance for MPQ algorithm, we assume that packets arrive at a queuing system according to a Poisson process of rate λ. Walked through the procedures above, the packets stored in the high priority queue are processed with the service rate of $A\mu$ prior to anything else. For this, weighting factor A is greater than 1. After there is no more packets for service left in the high priority queue, those in the low priority queue are provided for service in a distribution with the average of μ.

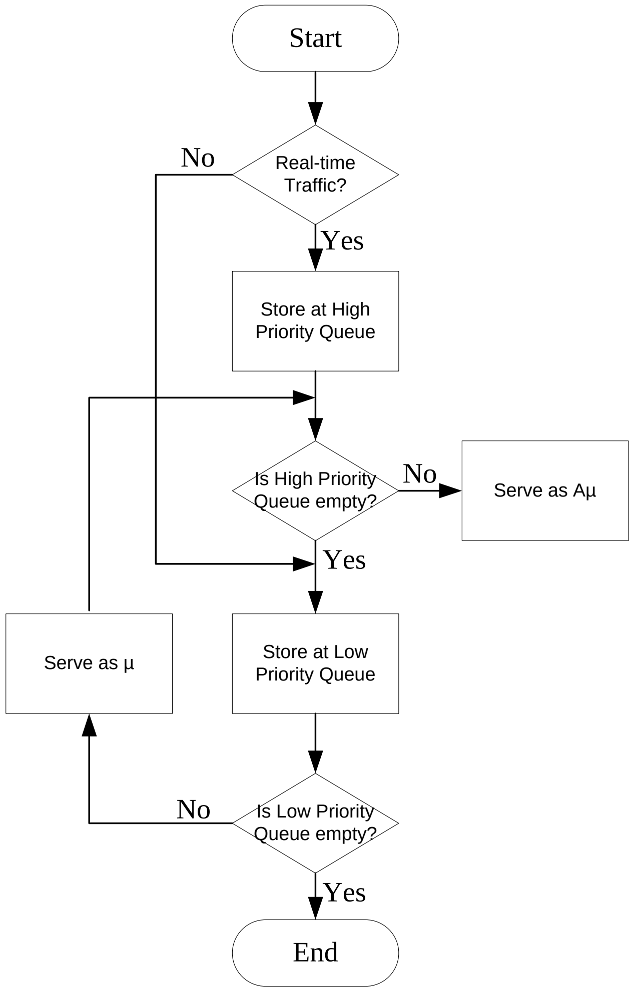

Fig. 1 The flowchart of MPQ Algorithm

As a result, as service of the input packets is provided on the basis of the priority in accordance with the traffic class, it is advantageous in meeting the delay QoS requirement due to the ability to reduce the delay on account of more and faster service from the high priority queue than the low priority queue. And weighting factor A can be efficiently chosen to consider the QoS requirement on the packets left in low priority queue that are not unprocessed.

Fig. 1 is detailed as follows. For the real-time traffic case, the traffic occurring with Poisson distribution arrives with the rate of λ and is stored in the high priority queue, and service is provided with $A\mu$ of exponential distribution until nothing is left in that queue. Mentioning the non-real-time traffic, the traffic occurring with Poisson distribution enters the low priority queue with the rate of λ and is serviced with μ of exponential distribution before the high priority queue gets to have packets. For the packet entering the high priority queue, service restarts to be provided with $A\mu$ of exponential distribution.

3 Numerical analysis

We consider MPQ a queueing system that handles K priority classes of users. Type k users arrive according to a Poisson process of rate λ_k and have service times with PDF $f_{\tau_k}(x)$ and mean $E(\tau_k)$. A separate queue is kept for each priority class, and each time the server becomes available it selects the next user from the highest-priority nonempty queue. This service discipline is often referred to as "head-of-line priority service". We assume that users cannot be preempted once their service has begun [4].

In this paper, we consider a simple queueing system that real-time traffic arrives with high priority according to a Poisson process and has exponentially distributed service times with mean $1/(A\mu)$ and non-real-time traffic with low priority arrives according to a Poisson process and has exponentially distributed service times with mean $1/\mu$. Suppose that high priority users are given non-prermptive priority. Therefore, we find the mean waiting time assuming two-priority M/G/1 system with priority given to real-time voice traffic. When we assume that the real-time and non-real-time traffic cases have the arrival rate into high priority and low priority λ_{high} and λ_{low} of Poisson distribution, respectively, the first two moments of the service time are [5]

$$E[\tau] = \frac{\lambda_{high}}{\lambda_{high}+\lambda_{low}} E[\tau_{high}] + \frac{\lambda_{low}}{\lambda_{high}+\lambda_{low}} E[\tau_{low}] = \left(\frac{\lambda_{high}}{\lambda_{high}+\lambda_{low}} \frac{1}{A\mu} + \frac{\lambda_{low}}{\lambda_{high}+\lambda_{low}} \frac{1}{\mu} \right) \quad (1)$$

$$E[\tau^2] = \frac{\lambda_{high}}{\lambda_{high}+\lambda_{low}} E[\tau_{high}{}^2] + \frac{\lambda_{low}}{\lambda_{high}+\lambda_{low}} E[\tau_{low}{}^2] = \frac{\lambda_{high}}{\lambda_{high}+\lambda_{low}} 2\left(\frac{1}{A\mu}\right)^2 + \frac{\lambda_{low}}{\lambda_{high}+\lambda_{low}} 2\left(\frac{1}{\mu}\right)^2 \quad (2)$$

The traffic intensity for each traffic class and the total traffic intensity are

$$\rho_{high} = \lambda_{high} E[\tau_{high}] = \lambda_{high} \frac{1}{A\mu}, \quad (3)$$

and $\rho_{low} = \lambda_{low} E[\tau_{low}] = \lambda_{low} \frac{1}{\mu}$

$$\rho_{total} = \lambda_{high} E[\tau_{high}] + \lambda_{low} E[\tau_{low}] = \lambda_{high} \frac{1}{A\mu} + \lambda_{low} \frac{1}{\mu} \quad (4)$$

where λ_{total} is the total arrival rate. The mean residual service time is then

$$E[R^*] = \frac{\lambda_{total} E[\tau^2]}{2} = \lambda_{high} \left(\frac{1}{A\mu}\right)^2 + \lambda_{low} \left(\frac{1}{\mu}\right)^2 \quad (5)$$

The mean waiting time for M/G/1 system with priority service discipline is

$$E[W_{high}] = \frac{E[R^*]}{1-\rho_{high}} = \frac{\lambda_{high}\left(\frac{1}{A\mu}\right)^2 + \lambda_{low}\left(\frac{1}{\mu}\right)^2}{1-\lambda_{high}\frac{1}{A\mu}} \quad (6)$$

$$E[W_{low}] = \frac{E[R^*]}{(1-\rho_{high})(1-\rho_{high}-\rho_{low})} = \frac{\lambda_{high}\left(\frac{1}{A\mu}\right)^2 + \lambda_{low}\left(\frac{1}{\mu}\right)^2}{\left(1-\lambda_{high}\frac{1}{A\mu}\right)\left(1-\lambda_{high}\frac{1}{A\mu}-\lambda_{low}\frac{1}{\mu}\right)} \quad (7)$$

where $E[W_{high}]$ is the mean waiting time for high priority user and $E[W_{low}]$ is the mean waiting time for low priority user.

4 Results of Simulation

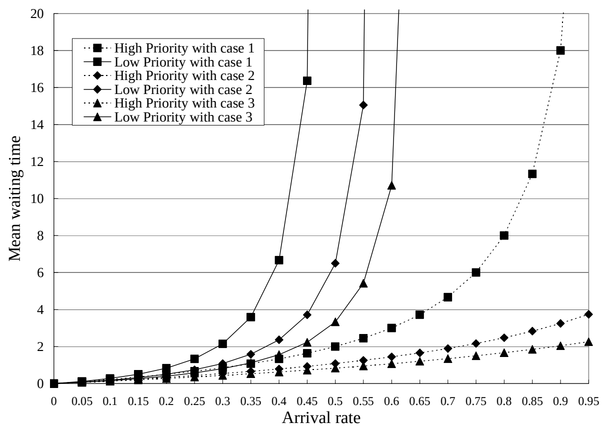

Fig.2 Mean waiting time vs. arrival rate $(\lambda = \lambda_{high})$ with variable weighting factor A

Fig. 2 shows the per-class mean waiting time versus $\lambda_{high} = \lambda_{low} = \lambda$ with different weighting factor A such as case 1 $(A=1)$, case 2 $(A=1.5)$ and case 3 $(A=2)$. It can be seen that the discipline "short-job type first" used here improves the average wait time. In case 3 with $A=2$, Fig. 2 shows that at λ=0.6 the lower-priority queue becomes unstable but the higher-priority queue remains stable up to λ=0.95. From the above result, we observe that case 3 has the less mean waiting time than case 1 and case2. Therefore, in MPQ, we find that it is effective that the larger the weighting factor assigned to the real-time service that is sensitive to the delay is, the smaller the mean waiting time is.

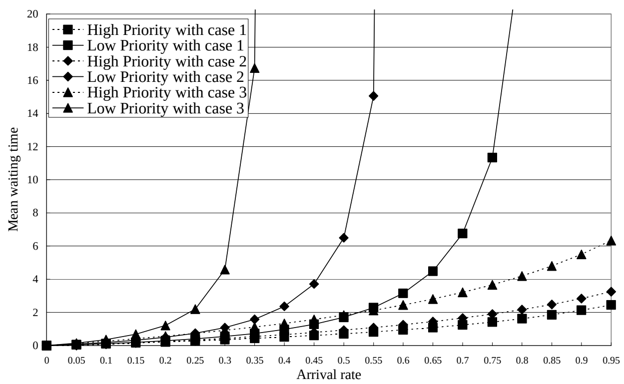

Fig. 3 Mean waiting time vs. arrival rate $(\lambda = \lambda_{high})$ with different λ_{high} and λ_{low}

We suppose that users arrive according to a Poisson process at a rate of case 1, case 2, and case 3, where case 1 $(\lambda_{high} = 0.5\lambda_{low})$, case 2 $(\lambda_{high} = \lambda_{low})$, and case 3 $(\lambda_{high} = 2\lambda_{low})$. Fig. 3 shows the mean waiting time versus the variable rate of λ_{high} and λ_{low}. For example, Fig. 3 shows that the lower-priority queue with case 1 becomes unstable at $\lambda = \lambda_{high} = 2\lambda_{low}$ =0.8 but the lower-priority queue with case 3 becomes unstable at $\lambda = \lambda_{high} = 0.5\lambda_{low}$ =0.35. In addition, for the higher-priority queue the same results are obtained. From the result of Fig. 3, we observe that case 1 has the less mean waiting time than cases 2 and 3. We know that as the arrival rate of real time service with high priority is faster than that of non-real time service with low high priority, MPQ algorithm does good in maintaining less mean waiting time.

5 Conclusion

Generally, for the real-time traffic case, the priority queuing (PQ) packet scheduling algorithm enables service to be provided on the basis of high priority, and when there being no high prioritized traffic, it goes to the low priority packet. However, proposed modified priority queuing (MPQ) packet scheduling algorithm, realizing the high and low priority traffic services employs exponential distributions with the averages of $A\mu$ and μ respectively, secures the reduction in the mean waiting time regarding the high priority traffic sensitive to delay. Thus, by adjusting the weighting factor A, we could meet the requirements for the mean waiting time of the real and non-real-time traffic services. As a result, the proposed MPQ packet scheduling makes difference by applying higher service rate to the traffic class with higher priority than that with lower priority. The proposed MPQ packet scheduling algorithm can be usefully used in wireless communication system using adaptive modulation and coding (AMC) due to time-varying capacity by providing the traffic class with higher priority with the higher order modulation with higher code rates.

References:

[1] 3GPP TSGR1 #16(00) 1316, "TR on HSDPA", 16 October 2000.

[2] 3GPP TSGR1 #12(00) 0556, "Feasibility study of Advanced techniques for High Speed Downlink Packet Access", Motorola, 10-13th. , April, 2000.

[3] Geoff Huston, *Internet Performance Survival Guide*: QoS Strategies for Multiservice Networks, Wiley Computer Publishing, 2000.

[4] Leonard Kleinrock, *Queueing Systems volume 1: Theory*, 1975.

[5] Alberto Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Second Edition, Addison-Wesley Publishing Company, 1994.