

Observations on using marks for pricing in multiclass packet networks to provide multidimensional QoS

Juan Alvarez¹ and Bruce Hajek²

Department of Electrical and
Computer Engineering and
the Coordinated Science
Laboratory

University of Illinois
Urbana, IL 61801

e-mail:

{alvarez,b-hajek}@uiuc.edu

Abstract — We investigate through simple simulations some basic issues surrounding the use of marks for pricing in multiclass packet networks with multidimensional quality of service(QoS) profiles. We find that the use of only one class of packets is not sufficient to provide satisfactory multidimensional QoS performance in some simple situations. We also find that by using appropriate control mechanisms (e.g. pricing, marking and scheduling) and multiple packet classes it is possible to satisfy different QoS profiles within the same network. And last, we present some trade-off relationships among different QoS parameters (e.g. throughput, loss) that arise in a multiclass packet system. We assume both pricing based on packet marks imposed by the network (as suggested by Gibbens and Kelly and a small number of packet classes (as proposed by Clark), to achieve end-to-end congestion control and multidimensional QoS. Different marking and scheduling schemes are used to accomplish the goal of providing multidimensional QoS while maintaining efficient use of the network's resources.

I. INTRODUCTION

There are two important issues that we observe in the Internet today. One of them is that it does not provide differentiation between packets. Therefore, all users (or applications) have to accept the best-effort guarantees that currently exist, no matter what their actual quality of service(QoS) requirements are, even though it is clear that the Internet supports a mixture of users with different criteria for network performance (QoS profile) as pointed out by Clark [3]. The other issue is the increase in aggressiveness of TCP, as well as the increase in the use of non-congestion-controlled best-effort traffic. This could lead to “congestion collapse”, as indicated by Floyd and Fall [4]; unless end-to-end congestion control is employed by the users, and incentives to use it are deployed by the network. We investigate a solution to both of these problems using congestion-based pricing with multiple packet

classes by combining and extending the ideas of Gibbens and Kelly [1], Clark [3] and MacKie-Mason and Varian [6].

Clark [3] proposes to allow the classification of packets into service classes in order to provide different service levels to different users to meet their different needs. Let's suppose that a user only cares about the time it takes to transfer an object (a web page, a file, etc.) as the measure of network performance. Then, since the user knows the sending rate it requires in order to achieve its goal, it can set it as its service profile. It can then use a traffic meter to mark packets as *in* or *out* of its profile. *Out* packets are more likely to be dropped in case of congestion. This case illustrates the use of two classes of packets to receive a desired QoS, although it is not guaranteed that their profile will be met, since resources might not be sufficient. Notice that even though there are two classes of packets, the QoS profile is one dimensional (i.e. it specifies the user's desired sending rate).

Gibbens and Kelly [1] explore the implementation of network pricing by charging users for marks that routers place on packets in order to achieve congestion control. There has been a drive to include Explicit Congestion Notification(ECN) bits in TCP [5], and there is an ongoing debate regarding the number of ECN bits and the mechanisms that should be used to set them. The ECN bits could be used as the marks needed for the pricing scheme just mentioned. This pricing scheme allows users that are willing to tolerate more marks per unit time to get a higher share of throughput. Again, we see that the user's QoS profile is one dimensional (i.e. it specifies the user's desired throughput share).

An interesting way to look at the problem is the “smart market” approach that MacKie-Mason and Varian describe in [6]. In this case, each packet carries a bid indicating the price the user is willing to pay for a packet, and the higher the bid the less likely it is to be dropped.

The use of utility functions and game theory has been shown to be useful in designing the network. The goal is to allocate the network resources in order to maximize the aggregate utility of the network, which can be considered as the sum of the utility functions of the individual connections). Another goal is to allocate the resources in a fair manner. These issues are closely related. For example, a suitable selection of utility functions can drive the network to select proportionally fair allocations of bandwidth, or approximately min-max fair allocations of bandwidth. Kelly [2] presents this approach with parameters that users and the network can choose to achieve the aggregate utility maximization in a proportionally

¹This work was supported in part by the Fulbright Foreign Student Program, and by Mexico's CONACYT.

²This work was supported in part by the National Science Foundation under contract NSF ANR 9980544.

fair manner. We could consider this as an auction in which the users' parameters are their bids for throughput and the marks imposed by the network as the target prices.

Recall that the QoS measure in the pricing scheme proposed by Gibbens and Kelly [1] is throughput. Suppose that a new user with a different QoS measure comes along (it might desire low loss or low delay) who is willing to pay for it by receiving less throughput. Now we have described a multidimensional QoS profile, and we ask:

Is it possible for this user to have its QoS requirements satisfied if there is only one class of packets in the network (i.e. if the network does not differentiate between packets with different QoS requirements)? We will address this question in different contexts in the next two sections.

We investigate the necessity of having at least a small number of different packet classes (possibly as small as two) and the behavior of the network in their presence. In order to distinguish between these classes, users place class labels on their packets and the network routers use the class labels to help make scheduling decisions. The packets labeled as high priority can be thought of as carrying a higher bid, and the marks they receive will have a higher price too. This leads to service differentiation. At this point we ask:

Is it possible to satisfy the different QoS profiles by using appropriate control mechanisms? and if so,

What tradeoff relationships exist among the different dimensions of the QoS profile (i.e. throughput, delay, loss, etc.)? We will also address these questions in the contexts described in Sections 2 and 3 by using pricing, marking and scheduling.

A competing approach, exemplified by ATM, RSVP and Frame Relay standards, is to use reservations and enforcement based on per-flow state information within network switches. It is important to notice that as the number of packet classes allowed increases, the closer the network comes to a per-flow controlled system.

In summary, in our approach the users communicate to the network routers by the rate at which they send packets as well by the class labels they place on their packets. The routers communicate to the users by the delay, loss and marks they impose on packets. The challenge is to use this two-way communication between users and network routers to allocate the network resources efficiently and provide multidimensional QoS.

The next two sections will show how a multiclass structure can be used to improve the utility of networks based on pricing. This utility takes into consideration the users' different types of QoS requirements. In Sections 2 and 3 we present two different contexts in which there are two types of users with differing QoS measures. The network provides service differentiation between classes and places marks on packets (service and marking mechanisms will be explained further along) in order to make efficient use of its resources. Users choose their expenditure rate (to "pay" for marks) and employ rate control algorithms in order prevent getting more marks than desired but also to reach their desired QoS. Section 2 presents a bufferless model of a network resource and Section 3 presents a resource with a buffer.

II. BUFFERLESS MODEL

We begin by describing Scenario 1 of Gibbens and Kelly [1] in order to provide a simple context. There are 20 "elastic users" (Gibbens and Kelly [1] describe a model for elastic

users) that transmit packets over a bufferless link in discrete time. The time slot duration is considered to be one time unit, and each packet takes one slot to transmit. The link can handle up to $N = 10$ packets per slot. If Y packets are transmitted by the users in a slot and $Y > 10$, then $Y - 10$ of the packets are dropped, and all Y packets are marked; but if $Y \leq 10$ then packets are neither dropped nor marked.

Suppose the i th elastic user is willing to pay for w_i marks per unit time. It uses a state variable $x_i(k)$ to determine the number of packets it transmits in slot k . Specifically, the user transmits $X_i(k)$ packets, where $X_i(k)$ is an integer near $x_i(k)$. For brevity we omit description of the rounding method used in [1] which makes the short-time average of the integer variable X_i close to that of the real variable x_i . The variable x_i is updated according to the state equation

$$x_i(k+1) = x_i(k) + \kappa(w_i - f_i(k)) \quad (1)$$

where $f_i(k)$ is the number of user i 's packets marked in slot k , and κ is a constant. A large value of κ produces faster convergence but also higher variance, while a small value produces the opposite in both criteria. Note that if $x_i(0) = 0$ then $x_i(k)/\kappa = w_i k - (f_i(1) + \dots + f_i(k))$, which is the amount the user is willing to spend over the first k slots minus the amount spent over the first k slots. So in the long run, user i pays for marks at its target rate, w_i , and hence we call w the expenditure rate.

Simulations of Gibbens and Kelly [1] show that the throughputs of the users monotonically increase with the rate of expenditure w_i , so we can consider the w_i 's as the users' bids for throughput. They also noticed that the lower rate users pay slightly more per unit throughput since their load is somewhat more bursty and hence harder to multiplex. We reproduced their scenario and obtained the same results. We noticed that users experienced a 3.32% packet loss on average.

At this point we pose a question:

If a new user joins the network but it cannot tolerate having more than 1% of its packets lost instead of the 3.32% loss experienced by the rest of the users, is the network able to satisfy the QoS requirements of the new user?

Here we establish packet loss as a second QoS measure that complements throughput, and we refer to this scenario as Scenario 2. Notice that under the conditions specified for this scenario, the only way that this new user can try to get a different performance from the resource is by changing its expenditure rate. Figure 1 shows the effect of this new user's expenditure rate w on its own throughput and loss, as well as on the rest of the users' loss and aggregate throughput when the aggregate expenditure rate for the rest of the users is 2.1 marks/slot. The individual expenditure rates range from 0.01 to 0.2. If we look at the lower right plot in Figure 1 we notice that when the expenditure rate for the loss sensitive user ranges from 0.0001 to 1 marks/slot its loss percentage remains above its desired 1%. Therefore, the answer to the question is that the new user simply cannot get the service it requires in this scenario.

If the user is allowed to change the circumstances then it might be able to have its QoS profile satisfied. We illustrate this by considering the following approach. The new user could lobby the network operators to reduce packet loss by putting marks on packets whenever the load exceeds 8 (rather than 10). The whole population of users would then back off, reducing the loss rate substantially. However, the total throughput would also be reduced, and if the other users are

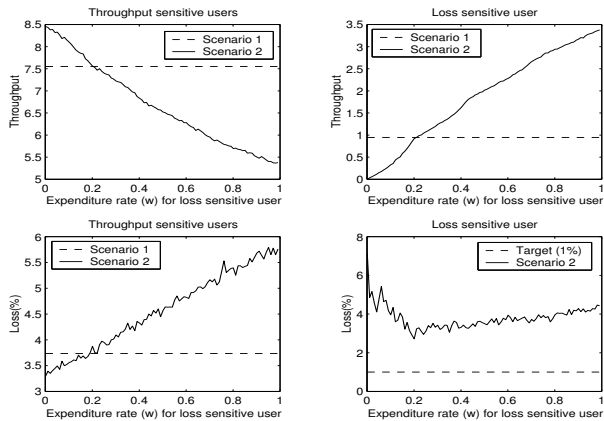


Figure 1: System behavior for Scenario 2.

not loss sensitive, this could lower the utility of the network (something that we are trying to avoid).

The solution we propose is to include a class label in the packets in order to differentiate between the two types of packets that we will refer to as *throughput sensitive* and *loss sensitive* packets. The router gives some preference to the loss sensitive packets and users pay a higher price for marks on packets that are labeled as loss sensitive.

We explore two scheduling mechanisms to give lower loss to loss sensitive packets:

1. *Capacity reservation*: reserve a portion of the capacity for loss sensitive packets and also reserve a portion of the capacity for throughput sensitive packets in order to prevent loss sensitive packets from starving throughput sensitive packets (i.e. the relation of loss to throughput sensitive reserved capacity can be 9:1,8:2,etc.).
2. *Strict priority*: a special case of capacity reservation in which the network gives preference to loss sensitive packets to use up capacity (i.e. the relation of loss to throughput sensitive reserved capacity becomes 10:0). Therefore, in case of congestion (i.e. $Y > 10$) service up to 10 loss sensitive packets. If the number of loss sensitive packets is less than 10, service enough throughput sensitive packets to reach capacity. Notice that even if there are throughput sensitive packets transmitted by users in a slot, possibly none of them could be serviced since loss sensitive packets are allowed to starve the capacity.

Consider the same setting as in Scenario 2, but now throughput sensitive users label all their packets as *throughput sensitive* packets, and the loss sensitive user labels all his packets as *loss sensitive* packets. The network makes use of the scheduling mechanisms mentioned above. Also, loss sensitive packets will have a higher price per mark. We will refer to this scenario as Scenario 3. We observe that by using either one of the two scheduling mechanisms and choosing an appropriate price for the loss sensitive packets it is possible to provide less than 1% loss to the new user, and hence satisfy its QoS profile. Figure 2 shows the system behavior using capacity reservation 1:9 (i.e. 1 space reserved for loss sensitive and 9 spaces reserved for throughput sensitive packets) as the scheduling mechanism.

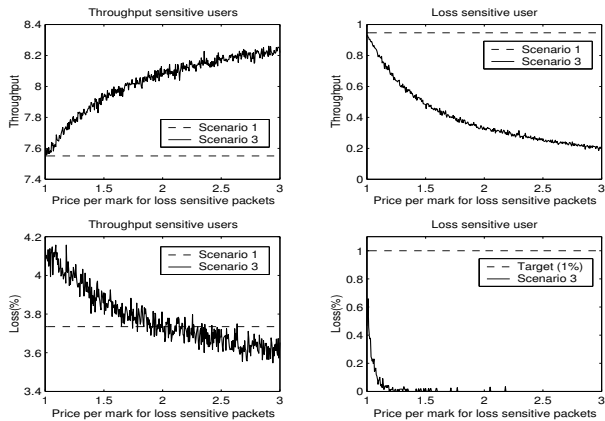


Figure 2: System behavior for Scenario 3.

It is obvious that the differentiation of the new user's packets must have some effect on the performance observed by the rest of the users, but maybe we want to be able to provide at least the same level of satisfaction to this users as in Scenario 1 as a fairness criteria. Therefore, the next logical question is:

Can the control mechanisms be adjusted so that the new user's QoS profile is satisfied while providing at least the same level of satisfaction to the rest of the users as in Scenario 1? Our simulations indicate that it is possible. We can see in the upper left plot of Figure 2 that the throughput for the throughput sensitive users is at least as good as before (in Scenario 1) if the charge per mark for loss sensitive packets is greater than 1. Also, the lower left plot of Figure 2 shows that the loss for the throughput sensitive users is at most what the experienced before if the charge per mark for loss sensitive packets is greater than 2.25. Therefore, if we set the charge per mark for loss sensitive packets to be greater than 1.75 we can provide at least the same level of satisfaction to the throughput sensitive users as before, so we'll call this operating region the 'fair region'. Notice from the lower right plot in Figure 2 that the loss for the loss sensitive user is zero in the fair region, so its QoS requirements are satisfied too. Since the loss for loss sensitive user is zero in the fair region, there is really no tradeoff curve in this case.

We will make a slight change of contexts for the sake of emphasizing these results. We will consider three scenarios similar to Scenarios 1, 2 and 3 but with the same number of throughput and loss sensitive users, as well as the same aggregate expenditure rates for both classes.

- Scenario 4: there are 40 throughput sensitive users sharing the bufferless link (with capacity to transmit 10 packets per slot). There is just one packet class so the network does not give preference to any packet over another. The aggregate expenditure rate is (4.2 marks/slot) with pairs of individual expenditure rates ranging from 0.01 to 0.2.
- Scenario 5: there are 20 throughput sensitive and 20 loss sensitive users sharing the same bufferless link. There is just one packet class so the network does not give preference to any packet over another. The aggregate expenditure rate for both types of users is the same (2.1 marks/slot) with individual expenditure rates ranging from 0.01 to 0.2 for throughput sensitive users. The aggregate expenditure rate, as well as the individ-

ual expenditure rates for loss sensitive users will be a parameter to be varied for comparison.

- Scenario 6: there are 20 throughput sensitive and 20 loss sensitive users sharing the same bufferless link. Throughput sensitive users label all their packets as throughput sensitive packets, and loss sensitive users label all their packets as *loss sensitive* packets. The network makes use of the scheduling mechanisms mentioned earlier in this section. The total expenditure rate for both types of users is the same 2.1 marks/slot) with individual expenditure rates ranging from 0.01 to 0.2 for both types of users.

Simulations show that in Scenario 4 users are able to get an aggregate throughput of 8.152 packets/slot and a 11.95% loss. However, suppose that the loss sensitive users desire a 1% loss.

Our simulations once again show that with a single packet class (i.e. Scenario 5), it is not possible to satisfy the loss sensitive users QoS profile, as shown in the lower right plot of Figure 3.

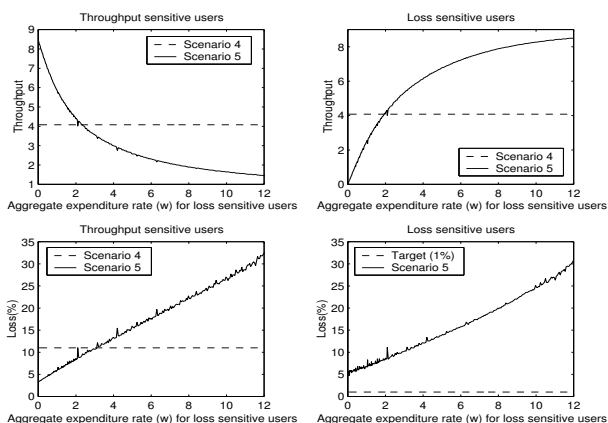


Figure 3: System behavior for Scenario 5.

In contrast, by using two packet classes and adjusting the control mechanisms in the network (i.e. price per mark, and scheduling), as in Scenario 6, it is possible to provide a 1% loss to the loss sensitive users at the expense of a reduced throughput for them. Figure 4 shows the system behavior using capacity reservations 6:4 (i.e. 6 spaces reserved for loss sensitive and 4 spaces reserved for throughput sensitive packets), 7:3 and 8:2.

Once again, it is possible to satisfy the QoS requirements for the loss sensitive users while providing at least the same level of satisfaction to the throughput sensitive users as before (in Scenario 4). This can be seen in the left side plots of Figure 4. The effect of using different capacity reservation choices on the tradeoff between loss and throughput for the loss sensitive users in the 'fair region' can be observed in Figure 5. This tradeoff curves can be helpful in deciding the appropriate capacity reservation to provide the required loss to the loss sensitive users while maximizing their throughput without deteriorating the performance seen by the throughput sensitive users. This can be done by choosing the curve with the highest throughput for the desired loss.

We tried different scenarios with higher expenditure rates, higher load, different ratios of *throughput/loss* sensitive users,

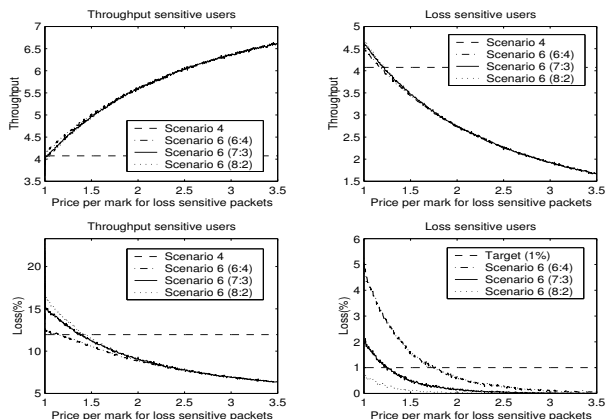


Figure 4: System behavior for Scenario 6.

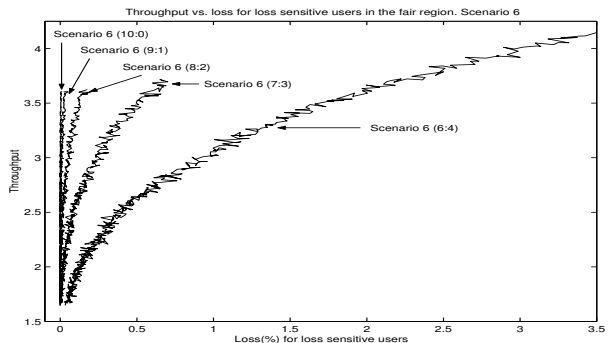


Figure 5: Trade-off curves for Scenario 6.

and we noticed that in most cases it is possible to adjust the control mechanisms to satisfy the different QoS profiles in an efficient manner.

In summary, in this bufferless context, the combination of control mechanisms and multiple packet classes can achieve the goal of providing multidimensional QoS while maintaining efficient use of the network's resources in most cases, while a single class system usually cannot.

As mentioned above, this example while simple is not the most realistic. First, in most situations added throughput can make up for a higher loss rate—for example a user could use forward error control, such as sending each packet twice. However, here we are just using loss as a second quality of service measure (to augment throughput) for the sake of illustration. One can ask the same question when there is queueing in routers and a new user comes along that requires a smaller delay than the system offers. Such a scenario is examined in the next section.

III. QUEUEING MODEL

We now consider a different context. There are 40 users transmitting packets over a link in discrete time. Once again, the time slot duration is considered to be one time unit, and all packet lengths are constant with transmission time equivalent to one time slot. The link has a queue with capacity to hold 10 packets while the link is transmitting another one (i.e. the system can hold 11 packets at once). Packets are dropped at the time of their arrival if the buffer is full.

Suppose that the 40 users are throughput sensitive and have an aggregate expenditure rate of 0.42 marks/slot with pairs of individual expenditure rates ranging from 0.001 to 0.02. In this scenario, which we will call Scenario 7, the aggregate throughput is 0.92 packets/slot, the loss is 3.9% and the delay is 5.91 slots(including the slot necessary to transmit the packet).

Now suppose that half of the users (with half of the expenditure rate) desire to have a 1% loss. We will refer to these users as *loss sensitive* users, and to the rest of the users as *throughput sensitive* users.

Is the network able to provide a 1% loss to the loss sensitive users? Figure 6 presents the behavior of the system in this scenario, which we will refer to as Scenario 8. If we look at the lower right plot in Figure 6 we can see that when the aggregate expenditure rate of the loss sensitive users ranges from 0 to 1.25 marks/slot, their loss remains above their desired 1%. Therefore, our simulation shows that if only one packet class is used in the system, it is not possible to satisfy the loss sensitive QoS profile.

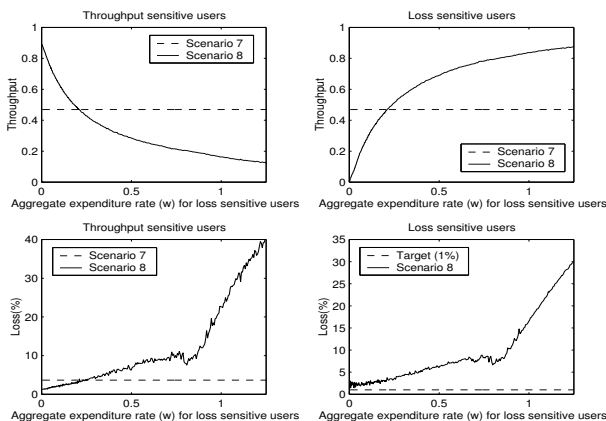


Figure 6: System behavior for Scenario 8.

We explore a solution to this problem by using two packet classes and a combination of scheduling and marking mechanisms described below.

We explore two scheduling mechanisms to provide low loss to *loss sensitive packets*.

1. *Strict priority* : if there is a loss sensitive packet in the queue service it, and service a throughput sensitive packet otherwise. Notice that if there is enough loss sensitive load, regular packets might never receive service.
2. *WFQ* : use weighted fair queueing [11] to service the two packet classes. The class with a higher weight will receive service more often. Notice that strict priority is the limiting case of WFQ obtained as the weight for the loss sensitive class approaches infinity.

We also explore two marking mechanisms:

1. Mark all packets between the occurrence of the first drop and the end of the corresponding busy period (as proposed by Gibbens and Kelly [1]). This is considered because if the queue size were a reversible stochastic process then the distribution of packet arrivals prior to the first drop is the same as that for the packet departures after the first drop (if such a packet drop exists in

the busy period). This marking mechanism is proposed with packet loss in mind.

2. Mark a packet if at the time of its departure from the queue the number of packets in queue is greater than the average delay experienced in the queue(this is similar to RED [11]). The reasoning behind it is that since packet transmission time is equivalent to one time slot, certainly that packet is causing a greater delay than average to other packets, and hence should be charged for it. This marking mechanism is proposed with delay in mind.

A third mechanism that can be used for the same purposes is *buffer reservation*. This mechanism guarantees part of the buffer for one class of packets, and the rest of the buffer is guaranteed for the other class. For example, if a 10:0 buffer reservation is used, then the whole queue is reserved for loss sensitive packets. However, the throughput sensitive packets can make use of this space when available.

Consider the same setup as in Scenario 8, but now the throughput sensitive users label all their packets as *throughput sensitive* packets, and loss sensitive users label all their packets as *loss sensitive* packets. The network gives preference to the loss sensitive packets by using the scheduling mechanisms described above and marks the packets according to the mechanisms also described above. Also, the price for marks on loss sensitive packets is higher than the unit charge for marks on throughput sensitive packets. We will call this Scenario 9.

Our simulations show that by adjusting the control mechanisms in the network (i.e. price per mark, scheduling, reservation and marking), as in Scenario 9, it is possible to provide low loss to the loss sensitive users. This is shown in the lower right plot of Figure 7 for three different combinations of scheduling and buffer reservation mechanisms using the loss oriented marking scheme.

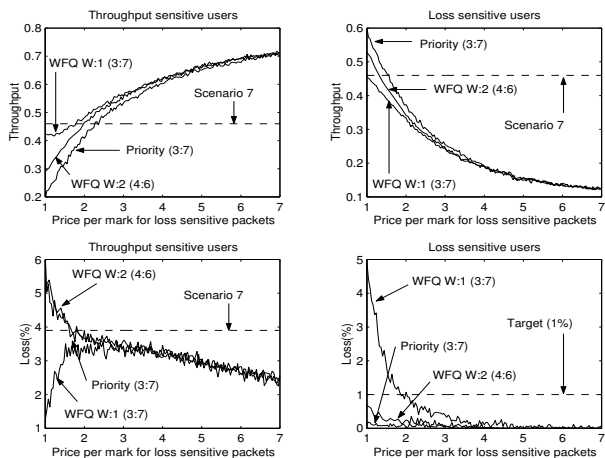


Figure 7: System behavior for Scenario 9.

In this context it is also possible to provide low loss to the loss sensitive while providing at least the same level of satisfaction to the throughput sensitive users, as shown by the left plots of Figure 7.

The effect of using different scheduling and buffer reservation mechanisms with different parameters on the tradeoff between loss and throughput for the loss sensitive users in the 'fair region' can be observed in Figure 5. This tradeoff

curves can be helpful in deciding the appropriate mechanisms to provide the required loss to the loss sensitive users while maximizing their throughput without deteriorating the performance seen by the throughput sensitive users.

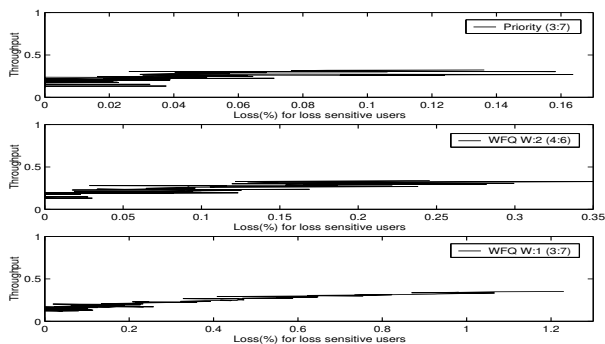


Figure 8: Trade-off curves for Scenario 9.

Again, we tried different scenarios with higher expenditure rates, higher load, different ratios of *throughput/loss* users, and we noticed that in most cases it is possible to adjust the control mechanisms to satisfy the different QoS profiles in an efficient manner.

In summary, in this queueing context, the combination of control mechanisms and multiple packet classes can achieve the goal of providing multidimensional QoS while maintaining efficient use of the network's resources in most cases, while a single class system usually cannot.

IV. CONCLUDING REMARKS

We showed through simulations that the use of pricing with only one class of packets is not sufficient to provide satisfactory multidimensional QoS performance in some situations. In contrast, the use of multiple packet classes in combination with pricing and appropriate control mechanisms (e.g. scheduling, marking, buffer allocation) can satisfy multidimensional QoS with efficient use of the network's resources within the same network. However, there are some trade-off relationships among different QoS parameters (e.g. throughput, loss, delay) that arise in a multiclass packet system.

The price per mark depends on the packet class and it should reflect the social cost required by the network to handle the packet. Since the QoS profile is multidimensional, the impact might be difficult to determine. We used two different marking schemes that made this evident. This paper only describes scenarios in which the loss oriented marking scheme was used. The reason for this is that we only considered loss as a QoS measure to augment throughput due to space limitations, but we have done similar work with the delay oriented marking mechanism. The main problem with the latter is that even if the average loss for loss sensitive users is below the desired level, the individual losses are not necessarily below the desired level.

An interesting question is how can we accommodate a multidimensional QoS profile within the notions of network wide fairness. Our fairness criteria was to provide at least the same level of satisfaction to throughput sensitive users when there are two packet classes as when there was only one class.

We described how users can receive better service than others in a particular QoS measure if they are willing to pay a

higher price per mark. Recall the "smart market" approach described by MacKie-Mason and Varian [6]. In auctions, bidders know the other bids and hence can hide the real price they are willing to pay (often called as shading bids), if there is not enough demand. After all, why pay more than enough? In the network case, could users "shade their bids" for service? Even if one user does not know the other users' bids, it could "shade its bid" if the service it has received so far is better than he expected (possibly if the load on the network is light, for example), since this performance is known to the user. Therefore, high priority users do not have to declare all their packets as high priority packets (they might only need to do it during congestion intervals) and hence, they can improve their overall service.

This paper considered two models for which loss and throughput are considered as two separate components of QoS. It remains to be seen if quality of service is fundamentally multidimensional for large scale networks and realistic traffic, or if instead, the dimensionality of the relevant QoS measures can be reduced to one in practice. This is a topic for further investigation.

Finally, we make the remark that a IETF DiffServ working group [10] is also addressing the issue of providing multidimensional QoS in the Internet.

REFERENCES

- [1] R.J. Gibbens and F.P. Kelly, "Resource pricing and the evolution of congestion control," Preprint, Statistical Laboratory, Cambridge, Latest version available at <http://www.statslab.cam.ac.uk/~frank/evol.html>, May 1999.
- [2] F.P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33-37, 1997.
- [3] D.D. Clark, "Adding service discrimination to the Internet," *Telecommunications Policy*, 20, 1996. Available at <http://ana-www.lcs.mit.edu/anaweb/abstracts/TPRC2-0.html>
- [4] S. Floyd and K. Fall, "Promoting the use of end-to-end congestion control in the Internet," *IEEE/ACM Transactions on Networking*, pp. 458-472, August 1999. Available at <http://www.aciri.org/floyd/end2end-paper.html>
- [5] S. Floyd's homepage for Explicit Congestion Notification in TCP/IP. <http://www.aciri.org/floyd/ecn.html>
- [6] J.K. MacKie-Mason and H.R. Varian, "Pricing the Internet," In B. Kahin and J. Keller, editors, *Public Access to the Internet*, Prentice-Hall, Englewood Cliffs, New Jersey, 1994. Available at <http://www.sims.berkeley.edu/~hal/people/hal/papers.html>
- [7] S. Shenker, "Fundamental design issues for the future Internet," *IEEE Selected Areas in Communications*, vol. 13, pp. 1176-1188, 1995.
- [8] F.P. Kelly, A.K. Maulloo, and D.K.H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237-252, 1997.
- [9] L. Massoulié and J. Roberts, "Bandwidth sharing: objectives and algorithms," *Proceedings IEEE INFOCOM '99*, March 1999.
- [10] IETF's webpage for Differentiated Services Working Group. <http://www.ietf.org/html.charters/diffserv-charter.html>
- [11] L.L. Peterson and B.S. Davie, *Computer networks: A systems approach*, Morgan Kaufman Publishers, Second edition, 2000.