

Exponential Scarcity: Metric For Fair Resource Distribution

Balázs Goldschmidt and Zoltán László
Department of Control Engineering and Information Technology
Budapest University of Technology and Economics

Abstract – *In online multimedia systems resources (e.g. network capacity) are to be evenly distributed among clients. These clients, however, usually have different QoS needs, and thus their actual capacity demands may vary considerably. In such context even distribution is not possible, but fair distribution, based on subjective client satisfaction must be achieved instead. In this paper a new metric, exponential scarcity, is proposed, that's based on clients' subjective quality experience. It is shown, that a resource distribution, where this metric is minimal leads to a balanced subjective satisfaction of the clients. As a corollary, the ideal scaling of QoS variations of multimedia content is proposed.*

Keywords: multimedia, QoS, fair service, metric

1 Introduction

As the internet access becomes more and more widespread, and as the available bandwidth and the capacity of home computers grow, the interest in multimedia contents, especially movies and music is steadily increasing [1, 2]. This growth of demand raises a lot of issues from the elaboration of appropriate formats to the question of copyright and intellectual property.

In the case of multimedia content delivery, the most resource demanding task is providing online and realtime playback, for this is what requires the most rigorous compliance to quality parameters. There is constant research in this area (Video on Demand) [3, 4], and distributed multimedia-servers offer the most promising results [5, 6]. These systems are, however, static in the sense that the components of the distributed server have to be deployed manually onto dedicated host computers.

The Adaptive Distributed Multimedia Server (ADMS) developed at Klagenfurt University tries to satisfy the clients' QoS demands with a different approach [7]. The server basically consists of three parts: (1) *Data Managers* (DM) that store the segments of the multimedia content, (2) *Data Collectors* (DC) that collect the segments from the DMs, and stream them to the clients, and (3) a central *Cluster Manager* that controls the connections of the two previous components and the clients.

In order to satisfy the quality demands as best as possible, the ADMS deploys the DCs in a proactive, dynamic way. Proactivity in this context means that – according to predicted future client-demands – the DC components are placed on appropriate points of the network so that the most clients can be served with the best quality possible.

The problem of how to measure and compare clients' satisfaction is a central question when the task is to improve client experience. There are two extremes of quality of services: in best effort QoS the client can not specify the quality, but may use the resources, and the system does its best to serve its clients. In hard QoS the system accepts or rejects clients depending on their quality needs. If a client is accepted, it is usually guaranteed that the quality will be provided. In multimedia services the clients usually define the range of quality they accept. Its lower bound is the worst acceptable quality, the upper bound is the presentation capability of the client's device. If the demand is accepted, the system tries to offer the client a quality in the range thus specified.

When handling multiple client requests, each with a quality range or quality list, a multimedia system has to choose a setting where the most clients can be served, and they get the best quality possible. During the elaboration of the cost function for the ADMS it was our concern to provide the fairest possible demand satisfaction, in other words, we wanted to serve the clients with equal quality, not preferring anyone to the others.

2 Related work

Measuring users' perception of multimedia quality, specified by QoS metrics is an important research area in the field of video on demand services. Past results try to help the optimization of users' satisfaction by the use of raw QoS parameters [8]. These results, however, mostly address perception regarding the defects of data flow, like packet loss and rate variations. In the context of ADMS our goal was to use such a metric that can be used to compare clients' satisfaction and to make resource distribution as fair as possible.

If the clients' satisfaction is to be taken into account, usually two options are used.[9]

- *Low level QoS measurement* where satisfaction is calculated by the measurement of connection parameters (bandwidth, jitter, latency, etc.), and subjective user experience is out of scope.
- *Subjective impact on clients* where satisfaction is evaluated by the clients themselves, and a formulaire is filled in, which is later used to create cumulated data.

The advantage of the first point of view is that in that case measurement is straightforward, proven techniques can be applied, and the results can be easily compared. The drawback is that usually clients do not have the knowledge to decide about the demanded values of low level parameters, and can get upset or confused if they have to adjust such things. Furthermore, the actual mapping of low level parameters to clients needs is not solved yet.

The second view eliminates the need to convert low level QoS data into the values of subjective experience, but rather handles these values directly. In that case, however, future demands can not be stated easily, because a correct mapping from subjective experience to low level QoS parameters would be needed.

Currently there are attempts of solving the problem economically, where measurable QoS levels are priced, and the clients decide which level has the best value/price ratio for them. For example in [10] an auction based pricing scheme is proposed for news-on-demand systems. In that case the client admission is driven by the search for an optimal profit/cost ratio. This approach, however, aims at renegotiated QoS for short time media access, which is out of our scope.

Our approach tries to give a simple mapping between subjective experience and low level QoS parameters, and tries to make it possible to make a fair distribution of resources by introducing the term 'exponential scarcity'.

3 Exponential scarcity

Let's assume that each client defines a demand list, where demand 0 describes the best quality, demand n the worse. We want to offer each client the demand with the lowest index possible. For each client let's call the index of the accepted demand the *accepted index*. First we defined linear scarcity, that was the sum of the accepted indices (for client c it is I_c):

$$S_{lin} = \sum_{\forall c \in C} I_c$$

Minimizing this metric we wanted to achieve the above goal. The problem is that this metric ignores the differences between the level of qualities each client gets. Lets consider an example setting, with three clients, the accepted indexes are:

$$I_1 = 2; \quad I_2 = 2; \quad I_3 = 2; \quad S_{lin} = 6$$

If we offer client 1 better quality and client 2 worse, then

$$I_1 = 1; \quad I_2 = 3; \quad I_3 = 2; \quad S_{lin} = 6$$

the linear scarcity is the same, while one client is preferred to the others. If the latter setting has less network demand, it is chosen. It's possible, that a few clients will get very good qualities, while the rest only much worse. This is why we call linear scarcity 'elitist'.

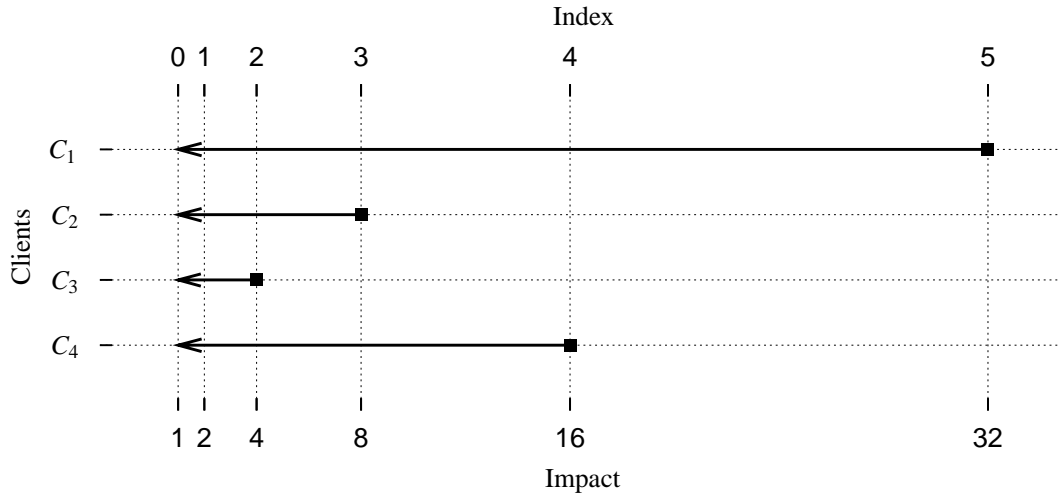


Figure 1: Example of exponential scarcity. On the figure each row represents the demand list of a client. For each demand at the top its index, at the bottom its scarcity weight are shown. On this example the accepted demands are marked with a box, their weights are shown as a wide arrow.

To avoid this phenomenon, we wanted to prefer clients with higher accepted indices (the ‘poor’) to those with lower indices (the ‘rich’). The metric became exponential

$$S_{exp} = \sum_{\forall c \in C} 2^{l_c}$$

This metric is better than the previous, because, if the resources allow, we get a smaller value when decreasing a higher acceptance index. Minimizing this metric we push the accepted indices from above. The higher the index, the greater its impact (see figure 1).

4 Detailed analysis

The above considerations assume that the clients have identical demand lists. In reality clients usually have very different demands, with different lower and upper bounds. In that case, how should a demand list be composed so, that for all clients the perceived subjective quality is almost the same?

In the analysis we use the following model. The resource demands are described by a single real value. Bigger value means more demand. The clients define an upper and a lower bound, the demand list is composed by dividing the range. The question is how this division should be made. The perceived subjective quality is the ratio between the upper bound and the accepted demand quality.

For client c let u_c be the upper bound, l_c the lower bound. The accepted index is l_c , the demand at this index d_c . For each following case we make a calculation, that shows the perceived subjective quality in function of the amount of available resource. The latter ranges from 0 to $\sum_c u_c$. The accepted indexes are chosen to give a minimum of exponential scarcity, while the total demand $\sum_c d_c$ is within the amount available resources.

1. Two clients, same bounds, different division.

Let’s have two clients that define the same bounds ($u_1 = u_2 = u$, $l_1 = l_2 = l$), but the second client has a shorter list.

Results show that the second client always gets worse quality than the first, except when there are enough resources available to serve the upper limits for both of them.

Conclusion: if the upper and lower bounds are the same, for equality the demand list has to have the same length.

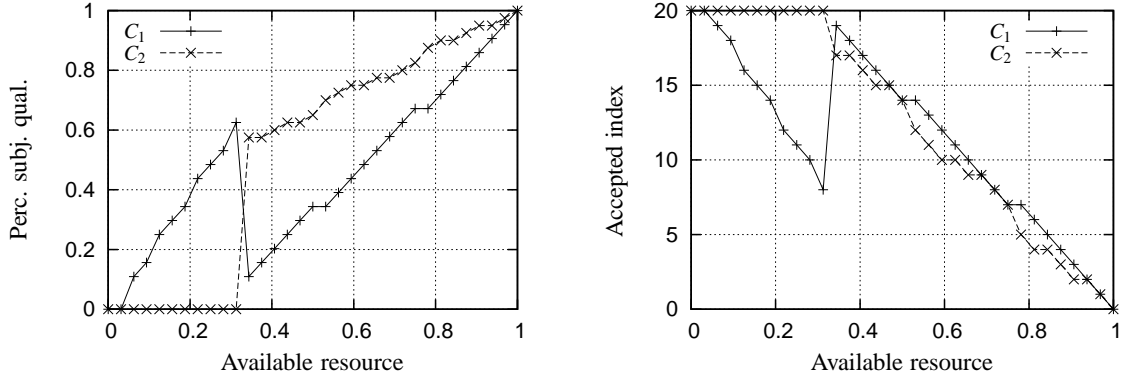


Figure 2: Example of two clients with same upper (16) but different lower bounds (1 and 8 respectively), and same lengths of demand lists (20). The demands are distributed linearly.

2. Two clients, same upper bounds, different lower bounds, same length of demand list.

Let's have two clients with the same upper bound ($u_1 = u_2 = u$), but different lower bounds ($l_1 < l_2$), and let their lists have the same length. An example setting can be seen on figure 2.

As long, as the available resources (r) are less, than the lower bound of the second client ($r < l_2$), the first client is served. When $r > l_2$, the second client is be served also, and, because the exponential scarcity is minimal if the accepted indexes are near equal, the second client gets better quality.

The above examples show, that although the clients' get different quality of service, their accepted indices are close. Based on that result we can calculate the optimal length of demand lists of the clients.

If the accepted range of quality is to be split exponentially¹, then the quality served to client c is (n_c is the length of the list):

$$d_c = u_c \left(\frac{l_c}{u_c} \right)^{\frac{l_c}{n_c}}$$

The perceived subjective quality is

$$p_c(I_c) = \frac{d_c}{u_c} = \left(\frac{l_c}{u_c} \right)^{\frac{l_c}{n_c}} = q_c^{\frac{l_c}{n_c}}$$

Clients with the same accepted indices should have the same perceived subjective quality, thus

$$\forall c, d : q_c^{\frac{l_c}{n_c}} = q_d^{\frac{l_d}{n_d}}; \quad \text{where } I_c = I_d$$

whence

$$q_c^{\frac{1}{n_c}} = q_d^{\frac{1}{n_d}} = C$$

$$n_c = \log_C q_c = \log_C \frac{l_c}{u_c}$$

Therefore the length of the list vary with the logarithm of the quotient of the lower and upper demand. C is an arbitrary constant, which impacts the length of the lists. An example can be seen on figure 3, which shows that the perceived subjective quality is indeed approximately the same.

Note, that the ratio between two consecutive demands in a list can be written as

¹The human senses usually perceive external signals with logarithmic scale (e.g. loudness, pitch, luminosity). Therefore it is recommended to use exponential scales in multimedia settings.

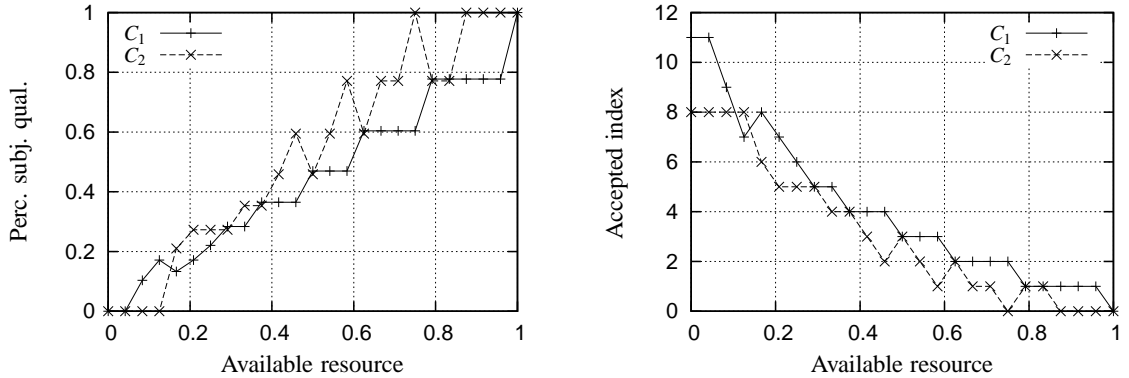


Figure 3: Example with five clients (only first two shown). Upper and lower bounds are $\{16,1\}$, $\{8,1\}$, $\{8,1\}$, and $\{20,2\}$ respectively. The length of the lists varies with the logarithm of the quotient of the upper and lower bounds, the distribution of the demands between the bounds is exponential.

$$\frac{d_{c,i}}{d_{c,i+1}} = \frac{u_c q_c^{\frac{i}{n_c}}}{u_c q_c^{\frac{i+1}{n_c}}} = q_c^{\frac{i-(i+1)}{n_c}} = q_c^{\frac{-1}{\log_c q_c}} = \frac{1}{q_c^{\frac{1}{\log_c q_c}}} = \frac{1}{C}$$

This result is the same for all clients, thus for all clients the ratio between consecutive demands is $\frac{1}{C}$. For every client the demands can be selected from the same initial set. If the demands describe the parameters of multimedia materials, then there is no need to create for each client an own variant of the material, they can use already existing materials instead.

5 Conclusion

If clients are to be served by Video on Demand servers, the server has to decide what QoS to provide to each client. In this paper we proposed a simple metric, *exponential scarcity* that makes it possible to provide fair distribution of resources among the clients. It is proven both theoretically and by simulations that optimization based on this metric leads to a solution where subjective client satisfaction is nearly the same even when resources are scare.

As a corollary, it was shown that there is no need to create for each client an own variant of the material, for every client the demands can be selected from the same initial set, and already existing materials are to be used only.

References

- [1] Mpaas homepage. <http://www.mpaas.org>.
- [2] Cds and dvds to go the way of the lp, September 2003. <http://www.forrester.com/ER/Press/Release/0,1769,831,00.html>.
- [3] Paolo Bellavista and Antonio Corradi. How to support internet-based distribution of video on demand to portable devices. In *Seventh International Symposium on Computers and Communications (ISCC'02)*, page 126, July 2002.
- [4] B. Qazzaz, R. Suppi, F. Cores, A. Ripoll, P. Hernandez, and E. Luque. Providing interactive video on demand services in distributed architecture. In *29th Euromicro Conference (EUROMICRO'03)*, page 215, September 2003.

- [5] Apple Computer, Inc. QuickTime Streaming Server. *Darwin Streaming Server: Administrator's Guide*, 2002. <http://developer.apple.com/darwin/projects/streaming/>.
- [6] Helix Community. *Helix Universal Server Administration Guide*, 2002. <https://www.helixcommunity.org/2002/intro/platform>.
- [7] Roland Tusch. *Design and Implementation of an Adaptive Distributed Multimedia Streaming Server*. PhD thesis, University Klagenfurt, Institute of Information Technology, April 2004.
- [8] Duminda Wijesekera, Jaideep Srivastava, Anil Nerode, and Mark Foresti. Experimental evaluation of loss perception in continuous media. *Multimedia Systems*, 7(6):486–499, 1999.
- [9] M. Ivanovich, P. Fitzpatrick, J. Li, M. Beresford, and A. Saliba. Measuring quality of service in an experimental wireless data network. In *Australian Telecommunications, Networks and Applications Conference (ATNAC)*, 2003.
- [10] Grzegorz Greg Malewicz and Alexander A. Shvartsman. An auction-based flexible pricing scheme for renegotiated qos connections and its evaluation. In *MASCOTS*, pages 244–251, 1999.