Multicast Extension of Unicast Charging for QoS Services

Antonio Bueno, Pere Vila, Ramon Fabregat
Broadband Communications and Distributed Systems Group
Institut d’Informàtica i Aplicacions (IIiA)
Universitat de Girona (Spain)
{bueno,perev,ramon}@eia.udg.es

Abstract

In order to successfully deploy multicast services in QoS-aware networks, pricing architectures must take into account the particular characteristics of multicast sessions. With this objective, we propose a charging scheme for QoS multicast services, assuming that the unicast cost of each interconnecting link is determined and that such cost is expressed in terms of quality of service (QoS) parameters. Our scheme allows determining the cost distribution of a multicast session along a Cost Distribution Tree (CDT), and basing such distribution in those pre-existing unicast cost functions. The paper discusses in detail the main characteristics of the problem in a realistic inter-domain scenario and how the proposed scheme would contribute to its solution.

1. Introduction

Pricing Quality of Service (QoS) services is a complex issue by itself. It can be divided in three different subtasks: accounting the information of the resource usage by the users; calculating the cost of this usage, i.e. the charging process we are interested in; and the billing and marketing processes which translate through offers, subscriptions and tariffs the cost plus the desired Network Provider (NP) benefits to the end-users (see figure 1). All these three tasks their selves are very complex and are very interrelated each other, however this paper focus only on the charging process proposing a mechanism as independent as possible from the other tasks. The aim of our charging scheme is to serve as basis to later determine the Service Provider (SP) pricing decisions. It is not the final price to be offered to end users, but rather a reference point that takes into account and integrates pre-existing unicast charging strategies.

If we consider the problem of multicast services, additional problems like fairness and usage sensitivity arise. Each participant in a multicast QoS service expects to receive the service at the agreed QoS, while being charged with a predictable amount, never higher than the amount applicable to an equivalent unicast service.

Therefore the considered scenario where the charging scheme is proposed and analyzed is as follows (see Figure 2). All nodes in the network are considered to be potential content distributors to end users and we call them Service Providers (SPs). One of these SP act also as content provider becoming the root of the Content Distribution Tree (CDT). Also note that these SPs can join and leave the CDT dynamically for many reasons, therefore this is a dynamic scenario.

Every SP can have its own complex internal network. Other scenarios with nodes in the content distribution tree that belong to the same SP or NP are included in this general case. It happens the same if there are nodes in the CDT that do not offer the service to end users. Therefore another interesting characteristic of this scenario is that it also includes in a natural way the fact that the multicast service with QoS guarantees is to be offered across several network domains, each represented by a node in our CDT. Summarizing, every node is potentially interested in receiving the service to sell it retail but also can be serving as a relay point to another domain further down the tree. The proposed multicast charging scheme aims to assign the portion of the cost attributable to each node, reflecting with it the

---

1 Article supported by Spanish Science and Technology Ministry (TIC2003-05567) and the Generalitat de Catalunya’s research support program (SGR 00296)
resources required to carry the traffic to such node, complying with the service QoS guarantees.

While along the paper we will refer to such multi-SP scenario, our charging scheme can be also interesting for a single SP case if it manages several subnetworks and wants to take into account the particular contribution of each subnetwork when offering a service across them. In fact this scenario and the original one are the extreme cases of the general problem of partitioning the cost between several entities collaborating to jointly offer multicast services with QoS guarantees.

In this paper we present a charging scheme for multicast services, with emphasis on their QoS characteristics. The target of this charging scheme is a multicast service deployed over a QoS-aware network, and it is especially suitable for the case where such multicast service is offered across several service providers (SP) that not only sell retail but also serve as multicast relay.

The proposed multicast charging scheme takes the unicast charging scheme between each SP (the nodes in our CDT) pair, and offers a cost partition that naturally extends the existing unicast charging agreements to the multicast case. The way the scheme is designed, it allows offering this cost partition in a fairly wide range of cases, and the cost assigned to each participant can closely fit the used resources, which is especially important in the multicast case due to its intensive resource use. Moreover it also contemplates protection mechanisms against the variability of a dynamic multicast session.

The field of IP pricing has been studied in the last decade, ranging from purely technical discussions on collecting the necessary information to sophisticated economical models. While it can be seen as mainly a question of studying the traffic, collecting information, processing it to produce a measure of the impact of the service in the network and using such measure to determine the tariff to apply to the end-user, the details can become overwhelming if we add to the problem QoS considerations. In the late nineties, charging in ATM networks was studied by European projects like CANCANN [1] and CASHMAN, while further study for both ATM and IP networks were carried out by the European project SUSIE [2].

For its own, multicast services are a particularly complex task, mainly due to the commented problems of high resource utilization, dynamic sessions. Pricing of multicast services it’s not a new problem. The work of Chuang and Sirbu [3] is usually referenced by their empirical but widely observed estimation of how many resources are saved by using a multicast tree instead of a whole set of unicast connections, but let us not forget this estimation was the basis for a charging scheme that related the cost assigned to each node to the utilized resources.

The earliest proposal for cost partitioning is from Herzog et al. [4]. They offered interesting ideas for sharing the cost along the CDT, and they have become a basic reference in multicast pricing. Other work on charging schemes for multicast pricing are Einsiedler’s [5], based in link weights with some DiffServ considerations, and Henderson’s [6], a protocol-independent proposal. Other than Hurley’s, all this work has not considered the case of QoS services.

Other related work has focused not in charging schemes. For example, Breitgand et al. [7] proposed a protocol-independent group size accounting mechanism that significantly reduced the network load of the accounting module.

While undoubtedly any charging scheme proposal has to take into account the network technology in use, as well as the underlying accounting module, our proposal is expected to inherit these particular traits from the unicast charging schemes that are used as
basis. On the other side, we will offer some insight of the possibilities of our charging scheme for the above pricing module, but leaving aside end-user marketing considerations. Our focus is in the branches of the tree, not the leaves.

The paper is organized as follows: after this introduction, section 2 discusses the particular issues of the joint charging/QoS/multicast problem. Section 3 formalizes the scheme, explaining in detail its components and their implications in the commented scenario, as well as the potential issues. Section 4 analyzes how our charging scheme tackles the problems discussed above. The paper ends with conclusions and future work at section 5.

2. Problem Specification

Main challenges in multicast charging come from two sides: By itself QoS multicast is complex and the addition of the charging adds new challenges. First of all, it has to be considered that even in the static case, where participants in a multicast session are known all along the duration of the service, determining the optimal multicast tree can be modeled as the Steiner problem in networks, which is NP-complete. Additionally, the multicast routing problem becomes more complex once we include the need to satisfy the QoS requirements of new services (see [8] for a good review of QoS multicast issues).

The usual view of a multicast session as a single, but complex, connection to a multicast group, can make the amount of resources involved seem large. But, leaving aside the management complexity, the needed resources are always lower than the joint amount of resources in the case where each receiver requires a unicast connection. On the other side, the amount of resources needed for QoS guaranteed forwarding can grow to unacceptable levels (either technically or economically) if the used QoS metric degrades significantly with each successive hop in the tree. For example, high requirements on packet loss or delay in the case of deep trees can render the service impossible to offer after a certain number of hops.

Another problem with multicast sessions is the possible heterogeneity between receivers (which remember, in our case includes the intermediate nodes). In the general case, this can be either the characteristics of their network or their particular QoS requirements. In our particular case, as we are initially contemplating only the case of an end-to-end QoS level associated to the service, only the characteristics of their network are of interest.

The most classical problems in multicast research are the consequence of the variability that comes with a dynamic session, where receivers can join and leave the session at will. Such operations must be carried on without service disruption and with the minimum of management cost. For example reconstructing the tree every time a receiver (a node of the tree, in our case a SP) joins or leaves the session can be intolerable in service disruption terms. A less aggressive method is use incremental changes through graft/prune mechanisms. But these changes will probably move the tree away from the optimal case.

<table>
<thead>
<tr>
<th>Step 1:</th>
<th>Source</th>
<th>A</th>
<th>B</th>
<th>Link</th>
<th>C wants to join the multicast session</th>
<th>Required resources due to B in A-B link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2:</th>
<th>Source</th>
<th>A</th>
<th>B</th>
<th>Link</th>
<th>Link</th>
<th>C</th>
<th>Required resources due to B in A-B link</th>
<th>Required extra resources due to C in A-B link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Forwarding may require additional resources

In the particular case of a multicast QoS service, join operations may require additional resources allocation (see figure 3). On the other hand, leave operations may either free resources (for example in the case of a peripheral node) or require reassigning the cost of the still used resources between the remaining receivers. The joining of a new node in the tree can always be processed as a leaf node at some tree level, however a specially problematic case is when a node that is not a leaf (i.e. a forwarding node) wants to leave the multicast session. It is important to note that the probability of leaving for a given node (SP) depends on many factors such as the number of end users it serves, the percentage of time in session in relation with the total session time, etc. and it is very complex to determine or model.

Charging brings its own set of problems to the mix. On one hand, economical considerations of recovering
deployment costs, maintenance costs and the marginal cost of each offered service. On the other hand, a charging scheme has to be usable, useful and acceptable to all the participants in the transaction.

One of the most interesting results of [1] was an exhaustive set of attributes to evaluate any charging scheme. Some of these criteria are how sensitive the charging scheme is to the real network resources used (usage sensitivity), whether the customer can get unexpected charges (fairness, predictability) or how flexible is for the network operator necessities (expandability, profitability, auditability). Some criteria can not possibly apply to the multicast case (e.g. distance independence).

In general, but especially in the multicast case, usage sensitivity and simplicity are somewhat opposites. This is because if the charge is closely related to the used resources, more knowledge of the network is required and more complex the charging process. In our inter-domain scenario, very different charging policies and the use of different network technologies make the problem more complex.

An essential ingredient of the problem is the distribution of the end-to-end quality requirements along the route. This NP-problem has been studied in the field of multicast routing, and it has been solved via \( \varepsilon \)-approximations for some function classes [10]. While it is an important part of the problem, the set of charging functions used in the links of a route is not expected to be very heterogeneous, and it is reasonable to consider only certain families of functions.

3. The Charging Scheme

As mentioned before, our working scenario consists in a multicast tree where each node represents a Service Provider (SP). Each SP is interested in the traffic, and some of them (B and D in figure 2b) retransmitting and even duplicating (B) it.

We assume each pair of SP has determined the cost of each one of their interconnecting links and that such costs are expressed in terms of quality of service (QoS) parameters. Our scheme allows determining the cost distribution of a multicast session along our tree, and basing such distribution in those pre-existing unicast cost functions.

Basing the multicast cost distribution in the unicast cost functions allows (up to a certain point) each SP to use whatever pricing model they prefer. The only limitation is to be able to apply the multicast service QoS parameters to all the cost functions along the path. This limitation depends heavily on the nature of the QoS metric: As mentioned before, a simple metric like bandwidth does not degrade with distance (number of hoops), while a certain packet loss is difficult to maintain end-to-end through many hoops. Other QoS metrics, like packet jitter or relative QoS, lay somewhat in the middle.

In this scenario we only intend to determine, under reasonable criteria, what part of the cost is each SP responsible for. It is our understanding that once a SP has the traffic, it makes the service available to their end users in a one-to-one basis (i.e. no more multicast traffic replication), and usually in a setup where QoS is not a problem.

Although simple pricing is a desirable trait for the end user and the network topology should be transparent (except maybe a distance factor) at the end user level, we consider such marketing/billing considerations out of the scope of our charging scheme. This does not mean our work has no interest for marketing/billing. On the contrary, it can serve as a baseline for the end-user service cost.

We call our scheme MXUCS, for “Multicast eXtension of Unicast Charging Schemes” (see [9] for a first version of it). MXUCS basically assigns each node the cost of the resources necessary to carry the traffic to them. In the tree example (see figure) the cost of the traffic from A to B with the service QoS parameters is assigned to B. C is assigned the cost for the traffic from B to C and the increase of QoS necessary in A to B to achieve the service QoS parameters from A to C.

Our multicast session \( M = (A, \Delta, Q_s) \) is composed of a source node (A) that is the origin of the offered service (\( S \)), a finite set of destination nodes (\( \Delta \)) and a set of end-to-end QoS requirements (\( Q_s \)) associated to the service. For every \( N \in \Delta \) two subsets of \( \Delta \) are of interest to us: \( P_N \) and \( B_N \). \( P_N \) is the set of nodes that precede \( N \) in its path and \( B_N \) is the set of nodes that share the same parent node as \( N \) (its “brothers”).

The formal formulation of the scheme is of a recursive form, and it is as follows:

\[
\text{Cost}(N) = \sum_{p \in P_N} \xi_{pq} \cdot \text{Cost}(p) + \sum_{p \in P_N, \forall j} \lambda_{pq} \cdot \delta_{pq}(Q_s, Q_p)
+ C_q(Q_s) \quad \forall N \in \Delta
\]

\( 0 \leq \xi_{ij} \leq 1 \quad \forall i, j \in \Delta \) are the parameters of the charging scheme and will be explained later. Each \( \delta_{ij} \) is the cost of the additional resources (\( Q_p \)) in the link that ends in node \( i \) required to carry the service up to node \( j \) respecting the service QoS requirements (\( Q_s \)).

\( C_i \) is the unicast cost function of the link that ends in node \( i \).
At first glance, the cost assigned to a node has three components: the first term deals with the problem of multicast resource savings, the second term with the maintenance of the end-to-end QoS requirements, and the third one with the cost of the last hop, which the node is fully responsible of.

The first term is a linear combination of the previously calculated cost of each link of the path from the source to node $N$. $\xi_{ij}$ serve as the parameters for the charging scheme, and determine how much, if any, of the shared resources cost is assigned to a node. We consider the case where $\xi_{ij}$ are zero or almost zero the most interesting variant of the scheme, as it is the closest to the ideal scheme for the usage sensitivity criterion, and the most appropriate for a competitive environment.

At the other end of the range of the variations available in our charging scheme is the case where all the $\xi_{ij}$ parameters are equal to 1. This case, although using multicast, would be virtually equivalent in distributed cost to the use of unicast connections. The receivers would be assigned the cost of all the resources required to carry the service to them, without any saved resources considerations from the use of multicast. While this case is far from the usage sensitivity criterion is may be of interest in non-competitive environments or where the variability and uncertainty of dynamic multicast sessions is a major concern.

While sub-optimal from the usage sensitivity point of view, a set of small $\xi_{ij}$ values may be enough in a competitive environment while offering some protection against variability along a dynamic session. In general the set of parameters to be used is expected to be agreed between all the participants, in order to avoid unfairness issues, but they are ultimately responsibility of the entities managing the link between the involved nodes.

The second term of the presented recursive formulation reflects the cost of the additional resources (see figure 3) necessary not to carry the service to a node, but to allow its further forwarding, coping with the potential degradation of the QoS parameters. The $\lambda_{ij}$ in the formula reflect that the cost of those additional resources is to be shared between all the sibling nodes. As with the $\xi_{ij}$ parameters, the more usage oriented version of the scheme requires that $\sum_i \lambda_{ij} = 1 \forall j \in \Delta$ (and equal to $\text{Card}(B_{ij})^{-1}$ in the homogeneous case), while the unicast-equivalent version requires that $\lambda_{ij} = 1 \forall i, j \in \Delta$

This second term depends heavily on the nature of the available QoS guarantees. If the service QoS requirements are expressed in a quantitative (non-numerical) QoS metric, the impact of this term is directly related to such QoS metric degradation with each hop along the multicast tree. Such degradation can render the service impossible to be offered more than a certain number of hops away from the source. This can be carried out either by a limit enforced by an admission control mechanism or assigning an unacceptably high cost. Although for simplicity reasons in this exposition we do not consider heterogeneous QoS requirements between the receivers (instead we work with a service QoS), it can be considered to offer a lower quality service to those nodes where the original service can not be offered because of the QoS metric degradation problem.

If the service QoS guarantees are of qualitative nature our charging scheme is still applicable, but the second term of the scheme formulation becomes more difficult to determine. In such scenario, the issue of how far down the tree can the service be offered may require some self-imposed constraint.

The third term is the simplest one, but it is this term where the main ingredient of MXUCS comes into play: the unicast charging functions between each node pair ($C_i$). The Service Providers specify how they value the traffic offered in each link of the multicast session through them, and with them our scheme extends the SPs criteria to the multicast case. Specifically the third term is the cost of the last segment of the path to the node $N$.

Note that this node is responsible for the full cost of that link, at the QoS requirements needed to carry the service up to the node. And this is even if there are nodes down the tree that also benefit from such traffic. Such downstream nodes may be responsible for part of that same cost (because of their corresponding $\xi_{ij}$ parameters), but this full cost ensures node $N$ will not have an unexpected increase in the assigned cost if those downstream nodes leave the multicast session in the middle of the service.

An important detail to observe is that the cost is for the QoS required up to $N$, nor further. The additional QoS requirements to redistribute the traffic are reflected in the second term of our multicast cost function for the downstream nodes. But even when nodes in other branches of the CDT induce a higher QoS level in the common path with $N$ (which in a globally optimized CDT would allow to reduce the QoS in the last link of the path to $N$), this third term still will assign the cost to carry the service to $N$, independently of the circumstances of the nodes in
those other branches. This setup results in a sub-optimal partition of the QoS requirements along the CDT, but much less sensitive to the variability of a multicast session.

Thanks to the unicast cost functions from the SPs, the distribution of the cost along the path to each node becomes the problem of finding an optimal distribution of the end-to-end QoS requirements along the links that form the path. As mentioned before, the general case of this problem has been studied in the field of multicast routing, and it has been solved for certain families of functions [10], but at this point we are only considering simple and homogeneous cost functions.

With the aim of illustrating how MXUCS works, we offer a numerical example comparing our scheme with two other schemes from Herzog’s et al. [4]: Equal Tree Split (ETS) and Equal Link Split Downstream (ELSD). We include also the unicast case as worst-case reference.

As service for our example we choose a high quality video stream, with numerical QoS requirements of packet loss probability not higher than $10^{-4}$ and jitter not higher than $10^{-1}$ seconds.

The MXUCS variation that is closer to Herzog’s schemes is when $\xi_{ij} = 0$ $\forall i, j \in \Delta$, that is, when the first term of our formula is null. For this same reason we will consider a homogeneous situation, where all the links use the same unicast charging function. This will result in a straightforward solution to the QoS partitioning problem.

The unicast cost function used in the example is the following:

$$C(l, j) = 10^{-4}\left(\frac{10^{-2}}{l} + \frac{1}{j}\right)$$  \hspace{1cm} (2)

where $l$ is the packet loss probability and $j$ is the jitter expressed in seconds. The function follows the intuitive notion that perfect quality has infinite cost and vice versa. The coefficients have been hand-picked to roughly fit the example prices in Walker’s charging proposal [1] for VoD (around 4€ per hour).

The results of applying the charging schemes to the CDT in figure 4 are shown graphically in figure 5 and numerically in table 1.

It can be observed how the equal distribution of the

<table>
<thead>
<tr>
<th>Node</th>
<th>Multicast cost</th>
<th>Unicast cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MXUCS</td>
<td>ETS</td>
</tr>
<tr>
<td>B</td>
<td>0.0011</td>
<td>0.0033</td>
</tr>
<tr>
<td>C</td>
<td>0.0026</td>
<td>0.0033</td>
</tr>
<tr>
<td>D</td>
<td>0.0026</td>
<td>0.0033</td>
</tr>
<tr>
<td>E</td>
<td>0.0026</td>
<td>0.0033</td>
</tr>
<tr>
<td>F</td>
<td>0.0055</td>
<td>0.0033</td>
</tr>
<tr>
<td>G</td>
<td>0.0044</td>
<td>0.0033</td>
</tr>
<tr>
<td>H</td>
<td>0.0044</td>
<td>0.0033</td>
</tr>
<tr>
<td>Total</td>
<td>0.0231</td>
<td>0.0231</td>
</tr>
</tbody>
</table>
total cost of ETS results in a heavy subsidizing of the far nodes by the nodes closer to the source. The deeper the CDT, the less fair is this cost distribution.

With ELSD node $C$ is significantly penalized for not having receivers downstream. On the other side, node $B$ is subsidized because all the nodes downstream from it, but also making it vulnerable to any change downstream. It can be seen in table 1 how the nodes that are two hops away from the source ($C$, $D$ and $E$) have the same cost assigned by our scheme.

Note also that either in MXUCS or in ELSD, node $F$ does not have any other node to share the cost of the additional quality in the link between $B$ and $D$ (unlike $G$ and $H$), so its assigned cost is higher. The difference is more pronounced in our scheme (see the exact amounts in table 1) because downstream nodes in ELSD are carrying part of the cost of every link in their path, while MXUCS only assigns the cost of the additional resources.

Finally, note in the table that the total multicast cost is the 53% of the unicast cost. Although we do not include in our work marketing/billing considerations, it can be interesting to observe that even assigning every node the maximum multicast cost of a node (0.0055 in MXUCS for node $F$) the total cost would be lower than the equivalent unicast case. This “node maximum-cost” could be used as a baseline to implement marketing/billing strategies.

4. Evaluation

The presented scheme is particularly good about the usage sensitivity criterion, as the scheme follows closely the used resources, or at least offers a purely resource-based solution when the first term of the formula is zero.

Another good point is its generality, because the use of the unicast charging functions brings the particular technology considerations of each SP into our multicast charging scheme. Although it is true that very different unicast functions may result in a significant difficulty of solving the QoS partitioning problem. Different implementations of QoS may also make compatibility become too complex.

The fairness of the scheme is also good, or at least as good as the fairness of the unicast schemes between each SP. This is because MXUCS targets specifically the used resources, and fair unicast agreements should naturally extend to our multicast case. The higher the first term of the formula is, the less fair the scheme becomes in terms of closeness to the used resources, as it assigns the cost of the shared resources more than once. The interest of this overcharging case is because of the protection against the variability of very dynamic sessions, which can be argued as another way of achieving fairness. Please note that fairness refers to charge similarly users that are in similar circumstances.

Besides overcharging the shared traffic, the scheme includes another mechanism to help against the variability of a dynamic session. The recursive assignation of costs without considering downstream nodes to determine the QoS requirements along the path protects nodes in one branch of the CDT against changes in another branch (except in the common part of their path). It also protects a node against changes downstream. This grade of independence comes with a price, as the QoS partitioning is not globally optimal. Nevertheless, a global optimum is ephemeral and costly when working with dynamic multicast sessions.

Despite this mechanism the weakest aspect of the scheme is its predictability. The predictability of the cost of a service (based primordially on previous experience) gets worse the more dynamic the service sessions are. It also gets worse when the scheme follows closely the used resources. Although the objective of our scheme is to serve as basis to later pricing decisions by the SPs, it is important to take into account results like the INDEX project [11], which clearly showed the significance of price stability and predictability for end-users.

5. Conclusions and Future Work

In this paper we have presented a charging scheme for multicast services with QoS guarantees. The presented scheme extends pre-existing unicast agreements (in the form of unicast cost functions based on QoS parameters) to offer a solution for the problem of cost distribution of a multicast service offered in an inter-domain scenario. The scheme offers a range of options for the Service Providers (SPs), and performs accordingly to the criteria of usage sensibility, generality and fairness, while offering a significant degree of protection against the variability of dynamic multicast sessions.

Future work includes exploring its applicability to specific scenarios, its interoperability with different technologies and its integration within a complete pricing architecture.
6. References


