Accommodating Heterogeneity in a Multicast Session Through a Receiver-based Data Replication Scheme.

Moufida Maimour

▶ **To cite this version:**

Moufida Maimour. Accommodating Heterogeneity in a Multicast Session Through a Receiver-based Data Replication Scheme.. [Research Report] LIP RR-2004-06, Laboratoire de l'informatique du parallélisme. 2004, 2+22p. hal-02102016

HAL Id: hal-02102016

https://hal-lara.archives-ouvertes.fr/hal-02102016

Submitted on 17 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Laboratoire de l'Informatique du Parallélisme

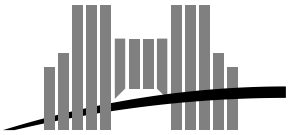
École Normale Supérieure de Lyon
Unité Mixte de Recherche CNRS-INRIA-ENS LYON n° 5668

Accommodating Heterogeneity in a Multicast Session Through a Receiver-based Data Replication Scheme

Moufida Maimour

January 2004

Research Report N° 2004-06



École Normale Supérieure de Lyon
46 Allée d'Italie, 69364 Lyon Cedex 07, France
Téléphone : +33(0)4.72.72.80.37
Télécopieur : +33(0)4.72.72.80.80
Adresse électronique : lip@ens-lyon.fr

Université Claude Bernard

Accommodating Heterogeneity in a Multicast Session Through a Receiver-based Data Replication Scheme

Moufida Maimour

January 2004

Abstract

A multicast session can involve multiple receivers with different capacities. To accommodate this heterogeneity, we propose a new replication mechanism which allows for a *fine-grained* multi-rate congestion control. In our scheme, some receivers (*replicators*) are responsible for data replication to a subset of receivers with lower capacity. A replicator, in the same way as a single-rate multicast source, adapts its rate depending on feedback it receives from the members of its associated subgroup. A simple partitioning algorithm is proposed to split a set of receivers into subgroups of similar capacities. This algorithm does not rely on a prior knowledge of the receivers' capacities and is executed on-the-fly as soon as necessary feedback are collected. To be more scalable and fairer with other sessions while improving the receivers satisfaction, we suggest to execute the partitioning algorithm at the routers. Analysis and simulations are performed in order to evaluate our approach, mainly by comparing it to traditional (source-based) replication schemes. Using ns, preliminary simulation results show the rapid convergence of the partitioning algorithm. Fairness of our scheme toward other flows is also dealt with.

Keywords: Reliable multicast, Congestion avoidance, Heterogeneity

Résumé

Une session multicast peut impliquer plusieurs récepteurs avec différentes capacités. Afin de prendre en charge cette hétérogénéité, un schéma de réplication par les récepteurs est proposé pour implémenter un protocole multi-débits, finement contrôlé. Dans cette approche, quelques récepteurs (*réplicateurs*) sont responsables de la réplication des données à un sous-ensemble de récepteurs avec une capacité inférieure. Un réplicateur, similairement à une source multicast à un seul débit, adapte son débit suivant les messages de contrôle qu'il reçoit des récepteurs membres de son sous-groupe associé. Un algorithme de partitionnement est fourni pour distribuer les récepteurs selon des sous-groupes de capacités similaires. Malgré sa simplicité, cet algorithme atteint ou au moins approche la solution optimale sans avoir besoin d'une connaissance préalable de la capacité des récepteurs. Afin de permettre plus de passage à l'échelle ainsi que plus d'équité avec les autres sessions tout en améliorant la satisfaction des récepteurs, nous proposons d'exécuter l'algorithme de partitionnement au niveau des routeurs. Nous montrons que notre approche de réplication par les récepteurs comparée à un multicast à un débit ou une réplication par la source, est plus équitable et permet plus de passage à l'échelle. Nous montrons également la convergence rapide de l'algorithme de partitionnement. L'équité avec les autres flux est également considérée dans ce travail.

Mots-clés: Multicast fiable, Contrôle de congestion, Hétérogénéité

Accommodating Heterogeneity in a Multicast Session Through a Receiver-based Data Replication Scheme *

Moufida Maimour.

1 Introduction

Multicast holds a great potential to drastically improve data delivery to multiple participants. It enables a wide variety of emerging applications, such as, software distribution, collaborative computing and multimedia conferencing. Since multiple nodes with different capacities could be involved in the same session, one of the challenging issues related to multicast, is how to accommodate receivers' heterogeneity. Single-rate multicast protocols [21, 15, 16, 6, 14] adjust their transmission rate in response to the most congested path in the multicast tree, which would limit the throughput of other receivers and thus their *satisfaction*. A multi-rate mechanism can improve the receivers' satisfaction, (known also as the *inter-receiver fairness* or equivalently the *intra-session fairness*), since receivers with different capacities can be served at a rate closer to their needs rather than having to match the speed of the slowest receiver. In a multi-rate session, the multicast source can transmit at different rates, either through a hierarchical scheme (layering) [18, 20, 5] or a replicated scheme (destination set grouping, DSG [10]). Layering schemes provide more economical bandwidth usage than DSG schemes, however layering is more complicated and requires efficient hierarchical encoding/decoding algorithms and synchronization among different layers.

A multi-rate multicast improves the intra-session fairness, however, fairness toward other unicast and multicast sessions known as *inter-session fairness* is required and has to be satisfied. Many works [2, 19, 9, 3] have addressed the concept of fairness, however earlier fairness definitions did not consider multi-rate schemes. More recently, Rubenstein et al. [17] studied fairness in the context of a multi-rate multicast. They identified four desirable properties of a *max-min fair allocation* and showed that a layered multi-rate scheme is more max-min fair than a single-rate scheme. However, Rubenstein et al. [17] did not consider replication-based multi-rate schemes in their study. In a DSG approach [10] (referred to as *source-based replication* scheme), data replication could be a source of unfairness. Allocated bandwidth to such a scheme at some common links, is almost the sum of all the rates of the replicated streams. This could make this replication scheme aggressive with other flows. We state that a multi-rate scheme (layered or replicated) could be a source of unfairness if it exists at least one link on which the consumed bandwidth is greater than the isolated rate of the fastest receiver among those located downstream this link. In a layered approach, there is no data replication and the cumulative rate of the different layers would not exceed the isolated rate of the fastest receiver. However, from a practical point of view, layered approaches fail to be totally max-min fair. The adaptation granularity is at the layer level since a receiver can not subscribe to fraction of a layer. To achieve a fine-grained adaptation, a layered scheme needs to have an infinite number of layers.

In this paper, we propose a replicated scheme where data replication is no longer the responsibility of the source. Some receivers (called *replicators*) contribute in the replication of the data flow with an appropriate rate to other receivers of lower capacity. In this way, a *regulation tree* is built with the source as the root, the replicators as intermediate nodes and the remaining receivers as final nodes. In order to minimize bandwidth consumption due to data replication, the regulation tree is built with respect to the physical multicast tree where routers are involved in the regulation tree construction process. Every router performs a partition of its downstream links into subgroups and chooses a replicator for every subgroup formed. In addition to the construction of a regulation tree with a topology close to the physical multicast one, executing the partitioning algorithm at the routers instead of the source is more scalable since, (i) every router performs

*Author may be reached via e-mail at Moufida.Maimour@ENS-Lyon.Fr.

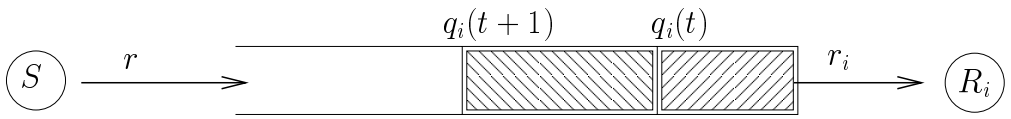


FIG. 1 – Queue length variation.

a partitioning algorithm locally, (*ii*) there is no data replication at the source which still send only one data flow, and (*iii*) the transmission rate of the source is no longer dictated by the worst receiver in the whole multicast group. Moreover, while inter-receiver fairness is improved, our approach satisfies inter-session fairness and allows for a fine-grained rate adaptation (congestion control). Independently of the source and other replicators, each replicator adjusts its rate according to the feedback it receives from its associated subgroup as a single-rate multicast source would do. The remainder of this paper is organized as follows. Section 2 provides our distributed algorithm for the regulation tree construction. An analysis of the proposed algorithm is provided in section 3. Afterwards a set of simulation results are presented in section 4. Fairness of our approach toward other sessions is dealt with in section 5 before concluding.

2 Regulation Tree Construction

The regulation tree construction is a distributed process where each router performs locally a partitioning of its downstream links into subgroups with similar capacities, and designates a replicator for every subgroup formed. Our partitioning algorithm is based on RTT measurements and is executed on-the-fly while the source is increasing its rate. There is no requirements on prior estimation of the receivers' (or links) capacities. In this work, the capacity of a receiver in a multicast session consists in its *isolated rate* [9] defined as the rate that a receiver would obtain if unconstrained by the other receivers in the group, assuming max-min link sharing. Earlier proposed partitioning algorithms [11, 22, 8, 4] are based on the prior knowledge of the receivers' isolated rates estimated based on RTT [8], loss rate [11] measurements or both of them [4]. Although simple, our algorithm approximates and in many cases, achieves the optimal partition without complex computations[1]. In [1], a computation is performed on every candidate solution before choosing the one that maximizes the receivers satisfaction. The dynamic programming algorithm proposed in [22, 8, 4] requires less computation effort but still be complex.

2.1 Preliminaries

In a multicast session, the satisfaction of a receiver R_i can be quantified using a utility function that maps the reception rate of the receiver to a normalized fairness value as the one proposed in [9] :

$$U_i(r) = \frac{\min(r_i, r)}{\max(r_i, r)} \tag{1}$$

where r_i and r are respectively the isolated rate and the R_i's reception rate. For what follows, we define the *RTT variation* noted $\Delta\tau$, experienced by a receiver as the difference between two consecutive RTT measurements. The *relative RTT variation* noted $\Delta\dot{\tau}_i$, is the ratio of the RTT variation to the amount of time elapsed between the two considered RTT measurements. If we assume that a receiver measures its RTT to the source every T seconds, then $\Delta\dot{\tau}_i = \Delta\tau/T$. In our approach, the partitioning algorithm is executed by the routers based on the RTT variations experienced by the links located downstream from them. The RTT variation of a router's downstream link is the RTT variation experienced by the worst receiver[2] among those located downstream from this link.

When the transmission rate (see Fig.1) exceeds the isolated rate of a receiver R_i, a queue of packets will build up within the path between the source and this receiver. We suppose that the receiver sends

[1] We do not consider here, the algorithm convergence to the optimal solution. For that, the interested reader can refer to [13].

[2] The worst receiver has the maximum RTT variation.

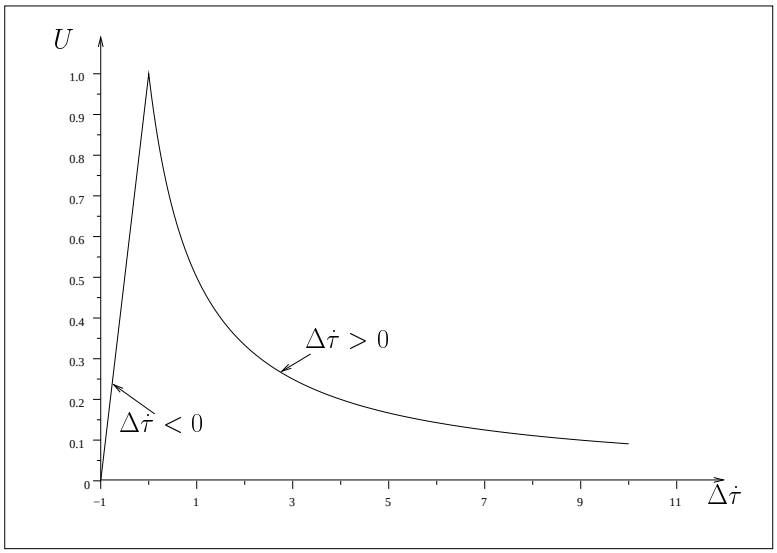


FIG. 2 – The inter-receiver fairness as a function of the relative RTT variation.

periodically a probing packet toward the source in order to estimate its RTT to the source. Let $q_i(t)$ be this queue size in packets at time t and Δq_i be the positive or negative variation in the queue length during a given time period T upon the reception of the subsequent probing. We have $\Delta q_i = q_i(t+1) - q_i(t)$ and noting by S the packets size, the queue builds up when $r > r_i$ during T with $\Delta q_i = (r - r_i)\ T/S$. We have $\Delta\tau_i = \Delta q_i\ S/r_i = T\ (r - r_i)/r_i$, which gives $\Delta\dot{\tau}_i = (r - r_i)/r_i$, then $r/r_i = 1 + \Delta\dot{\tau}_i$. Finally, the utility function (1) can be expressed as a function of the relative RTT variation (see Fig.2) as follows :

$$U_i(r) = \begin{cases} \dfrac{1}{1+\Delta\dot{\tau}_i} & \text{if } \Delta\dot{\tau}_i \geq 0 \\ 1 + \Delta\dot{\tau}_i & \text{if } \Delta\dot{\tau}_i \in [-1, 0[\end{cases} \tag{2}$$

A receiver that experiences a positive RTT variation could experience losses since its reception rate is greater than its capacity. In the case of a negative RTT variation, the receiver will be unsatisfied since it has more bandwidth resources. Note that the utility function is not defined for $\Delta\dot{\tau}_i < -1$ which corresponds[3] to a negative reception rate ($r < 0$).

In a similar way to [9], we define the utility of a multi-rate multicast session with receivers $\{R_1, R_2, \ldots, R_N\}$ split into K subgroups $\{G_0, G_1, \ldots, G_{K-1}\}$ with the transmission rates $g_0, g_1, \ldots, g_{K-1}$, as follows :

$$U(g_0, g_1, \ldots, g_{K-1}) = \sum_{k=1}^{K-1} \sum_{i=1}^{n_k} \alpha_{ik} U_{ik}(g_k) \tag{3}$$

subject to $\sum_{i,k} \alpha_{ik} = 1$ and $\alpha_{ik} \in [0, 1]$. We have $\sum_k n_k = N$ where n_k is the number of the receivers in subgroup G_k. $U_{ik}(g_k)$ and α_{ik} are respectively the utility function and the weight associated to the ith receiver of the kth subgroup.

2.2 Partitioning Algorithm

Each router performs a portioning of its downstream links among subgroups of similar capacities. Given the set of links $P_0 = \{l_j, j = 1, \ldots, B\}$ located downstream from a router, the problem consists in splitting this set of receivers into K subgroups (K can not exceed a maximum number G) to make a partition $P = \{P_0, P_1, \ldots, P_{K-1}\}$ of the original set P_0 so the overall session utility is maximized. We aim to determine

[3] see appendix A.

the optimal solution or at least an approximated one (without prior knowledge of the number of subgroups) such that the global utility is greater than a given threshold.

Algorithm 1 Regulation tree construction at a router

Require: $B > 1$ and $a < b$
  $P_0 \leftarrow \{l_j, j = 1, \ldots, B\}$, the set of all the links downstream
  $i \leftarrow 1$
  Periodically,
  if $\exists j, l_j \in P_0$ such that $\Delta\dot{\tau}_j > b$ **then**
    $P_i \leftarrow \{l_j \in P_0, \Delta\dot{\tau}_j > a\}$
    $P_0 \leftarrow P_0 - P_i$
    $Rep_i \leftarrow Best(P_0)$
    if $i > 1$ **then**
      $Rep_{i-1} \leftarrow Best(P_i)$
    end if
    $i \leftarrow i + 1$
  end if
  until $i = G$ or $|P_0| = 1$ or $\forall j < N, l_j \in P_0, \frac{1+\Delta\dot{\tau}_{j+1}}{1+\Delta\dot{\tau}_j} \geq \rho$

Initially (see algorithm 1), a router maintains the P_0 set with all the downstream links. Every time, a downstream link reports a relative RTT variation ($\Delta\dot{\tau}_j$) greater than parameter b, the router creates a new partition P_i with all the links that reported a relative RTT variation greater than parameter a ($a < b$) and selects a replicator Rep_i for this subgroup. The function $Best(P_i)$ returns for subgroup P_i, the identity of the receiver with the highest estimated capacity. When subgroup P_i is split from P_0, then $Best(P_0)$ is elected as its replicator. Since this replicator may have been chosen for P_{i-1} in the previous split, $Best(P_i)$ is definitely elected as the P_{i-1}'s replicator. The router continues spliting its links until G subgroups are already built or the P_0 is no longer "*split-able*", i.e. P_0 contains one element or remaining members are of similar capacities. The rationale behind these convergence criteria is provided in section 3.1.

During the algorithm execution as well as when the regulation tree is already built, a router forwards the source data packets only on the P_0 links. Every time, a new subgroup P_i is formed and the corresponding replicator Rep_i is selected, the router notifies this latter to start performing data replication. A replicator Rep_i sends its replicated data packets to the receivers of its corresponding subgroup P_i. Feedback messages from subgroup P_i are sent to their corresponding replicator Rep_i while those arriving on the P_0 links are forwarded to the source. Thus allowing for independent fine-grained congestion control at both the source and the replicators. Any single-rate congestion control algorithm could be used, and hence we do not rely on a specific congestion control algorithm to implement our approach. Due to space limitation, we do not consider here, issues on a practical implementation of our proposed approach. Nevertheless, the interested reader can refer to [12] that provides details about how a single rate congestion control protocol such as AMCA [14] could be extended using our approach to support more heterogeneous receivers.

2.3 Illustrative Examples

2.3.1 Example 1. "A star topology"

To illustrate our regulation tree construction algorithm, we consider a multicast session with seven subscribed receivers $\{R_0, R_1, R_2, R_3, R_4, R_5, R_6\}$ with respectively the following isolated rates $\{5, 6, 9, 11, 15, 19, 21\}$. All of these receivers are located downstream the same router A (see figure 3). If we take $a = 0.01$ and $b = 0.26$ which correspond to $\rho = 0.8$, executing algorithm 1 produces $\{\{R_0, R_1\}, \{R_2, R_3\}, \{R_4\}, \{R_5, R_6\}\}$ as a partition with the regulation tree shown in the right side of figure 3a. The resulting subgroups are $P_0 = \{R_5, R_6\}$, $P_1 = \{R_0, R_1\}$, $P_2 = \{R_2, R_3\}$, $P_3 = \{R_4\}$ with respectively the reception rates, $g_0 = 19$, $g_1 = 5$, $g_2 = 9$, $g_3 = 15$, which gives a session utility of 0.936 instead of 0.525 if no split is performed ($\forall i, \alpha_i = 1/7$).

The selected replicator for the first split subgroup $P_1 = \{R_0, R_1\}$ is $Best(P_0) = R_6$. In the next split, when the second subgroup P_2 is formed with $Best(P_0) = R_6$ as the replicator then the P_1 replicator (according to the proposed algorithm) has to be changed to $Best(P_2) = R_3$ which is definitely elected as a replicator for

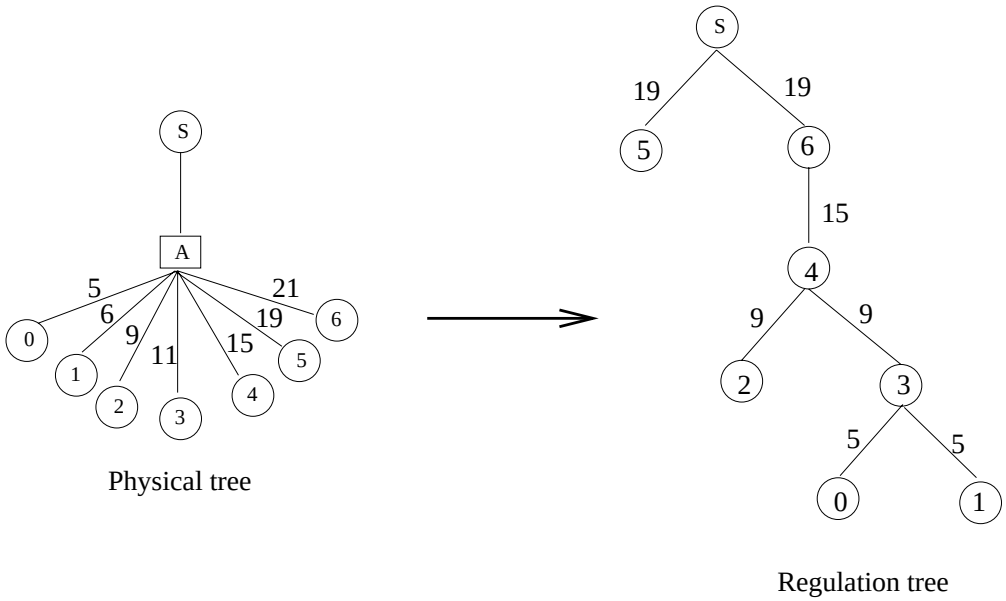


FIG. 3 – Example 1.

P_1. When $P_3 = \{R_4\}$ is split, its chosen replicator is $Best(P_0) = R_6$ and $Best(P_3) = R_4$ is definitely selected as a replicator for P_2 ... etc.

2.3.2 Example 2. "A hierarchy of routers"

In order to understand better our distributed partitioning mechanism, we consider one source multicasting to six receivers through three routers A_1, A_2 and A_3 (figure 4). A maximum bandwidth is given for every link in the considered multicast tree. The link (A_2,R_4) has a bandwidth of x which for the following will be set either to 10 or 2. The main concern here is the behavior of router A_1 since it has two indirect receivers R_3 and R_4 in addition to two direct ones, R_1 and R_2. In what follows, a link will be designated by its downstream node in the multicast tree. For example the A_1 downstream links that lead to router A_2 and receiver R_1 are respectively noted A_2 and R_1. We also use $P_{i,j}$ to designate the ith partition of router A_j.

Suppose that the maximum number of subgroups that could be built by a router is 2. If we set x to 10, then we get the following local partitions for the routers :

$$\begin{array}{rrrr} P_{0,1} = & \{R_1, A_2\}, & P_{1,1} = & \{R_2\} \\ P_{0,2} = & \{R_4\}, & P_{1,2} = & \{R_3\} \\ P_{0,3} = & \{R_6\}, & P_{1,3} = & \{R_5\} \end{array}$$

which gives the following overall partition seen by the source (figure 5a) :

$$P_0 = \{R_1, R_4, R_6\},\ P_1 = \{R_2\},\ P_2 = \{R_3\},\ P_3 = \{R_5\} \tag{4}$$

This partition (4) gives a utility value, $U_a = (1/1 + 3/3 + 5/5 + 10/10 + 10/11 + 10/12)/6 = 0.957$ instead of 0.301 if no partitioning is performed ($\forall i, \alpha_i = 1/6$). In this latter case the source would transmit data at a rate equal to 1 instead of 10.

If we consider that the partitioning is performed by the source instead of the routers, then the optimal partition with two subgroups is $P_0 = \{R_3, R_2, R_5\}$ and $P_1 = \{R_1, R_4, R_6\}$ which gives a utility value of $U'_a = 0.713 < U_a = 0.957$. In order to get a utility value equal to $U_a = 0.957$, a source-based partitioning requires four subgroups. In this case, the source link will be loaded with $10 + 1 + 3 + 5 = 19$ instead of 10 since the source will send four flows with rates 10, 1, 3 and 5. This gives $19/10 = 1.9$ of additional consumed bandwidth at the source link.

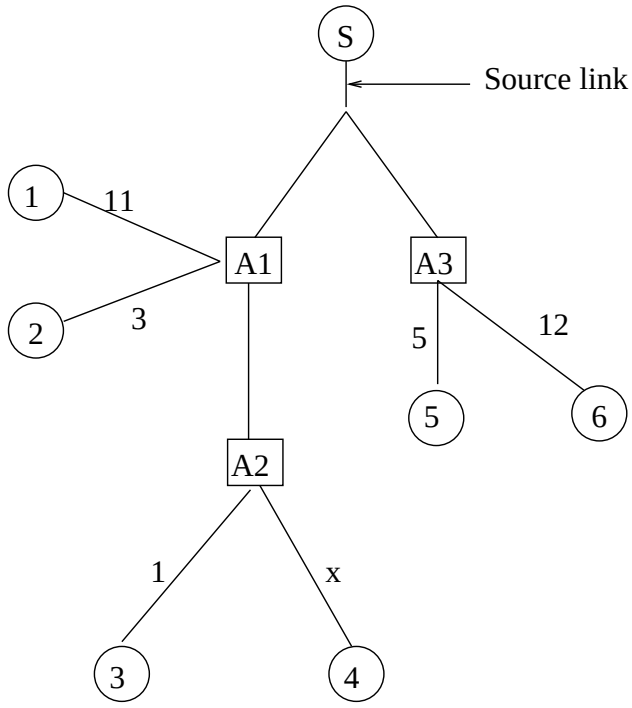


FIG. 4 – Example with a hierarchy of routers

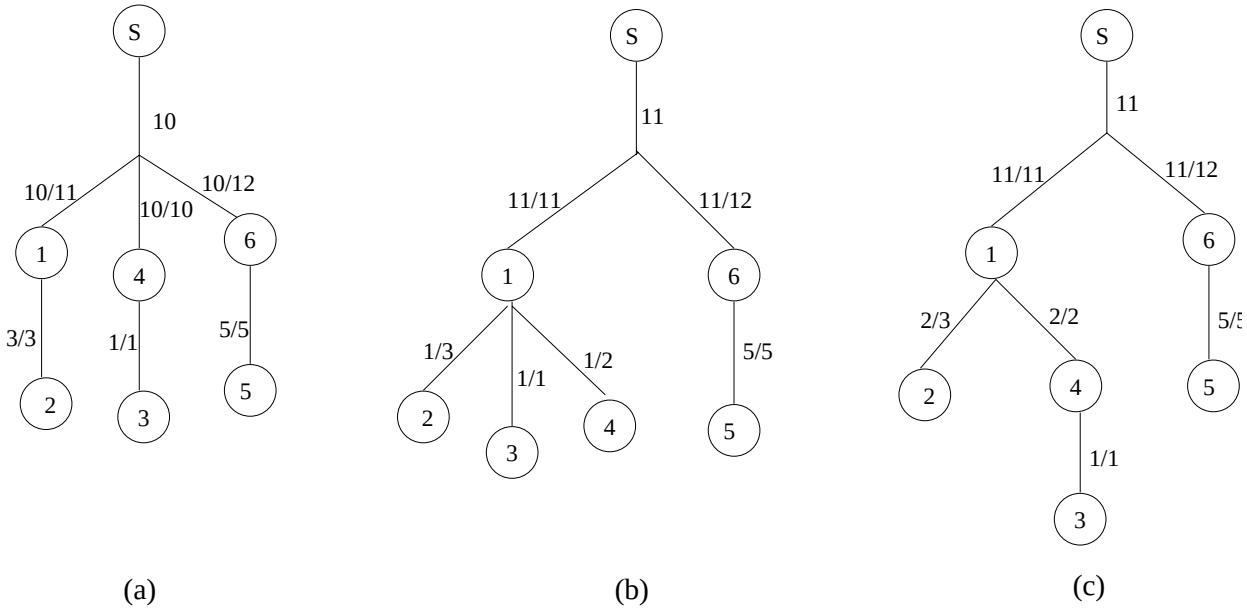


FIG. 5 – (a) $x = 10$, (b) $x = 2$, $\rho < 0.5$ and (c) $x = 2$, $\rho > 0.5$

Policy	Utility	Mean load	Load distrib.
Single-rate	0.368	1.0	0.0
Sender-based	0.667	8.1	27.01
Receiver-based	0.931	7.5	17.56

TAB. 1 – Example 2 : Utility and load distribution

If x is set to 2, then depending on ρ, we get respectively for $\rho < 0.5$ and $\rho > 0.5$:

$$\begin{aligned} P_{0,1} = &\quad \{R_1\},\ P_{1,1} = \{R_2, A_2\} \\ P_{0,2} = &\quad\quad \{R_3, R_4\} \\ P_{0,3} = &\quad \{R_6\},\ P_{1,3} = \{R_5\} \end{aligned}$$

and

$$\begin{aligned} P_{0,1} = &\quad \{R_1\},\ P_{1,1} = \{R_2, A_2\} \\ P_{0,2} = &\quad \{R_4\},\ P_{1,2} = \{R_3\} \\ P_{0,3} = &\quad \{R_6\},\ P_{1,3} = \{R_5\} \end{aligned}$$

that gives respectively the following overall partitions :

$$P_0 = \{R_1, R_6\}, P_1 = \{R_2, R_3, R_4\}, P_2 = \{R_5\} \tag{5}$$

and

$$P_0 = \{R_1, R_6\}, P_1 = \{R_2, R_4\}, P_2 = \{R_3\}, P_3 = \{R_5\} \tag{6}$$

When $x = 2$, the optimal source-based partition with two subgroups is $P_0 = \{R_2, R_4, R_3, R_5\}$ and $P_1 = \{R_1, R_6\}$ which gives a utility value of $U_b^{'} = 0.667$ instead of $U_b = 0.792$ or $U_c = 0.931$ for the partitions (5) (figure 5b) and (6) (figure 5c) respectively. In addition to the utility values, table 1 summarizes the mean load and its distribution among links in the different policies (single-rate, source-based and receiver-based replication schemes). We note that a receiver-based approach consumes less bandwidth than a source-based one. Moreover, a receiver-based replication allows for more load distribution among the different links. This is very important regarding the impact on fairness toward other flows as will be investigated in section 5.

3 Partitioning Algorithm Properties

To get an insight into our proposed algorithm, we consider the case of a tree topology with one level of routers (Fig.6). One source multicasts data to N receivers through M routers each of which has B downstream receivers. Let K be the number of subgroups built by every router, then the overall number of subgroups will be $M(K-1)+1$ instead of K if the partitioning algorithm is executed by the source. We can see in example 2 of the previous section, that a network topology with more hierarchy levels would produce a larger number of subgroups.

3.1 Star Topology

We first begin by analyzing the case of a star topology with only one router. In this case, the number of the receivers N is equal to the number of the router's downstream links B. Hereafter, we will use *"receiver"* and *"link"* interchangeably, since in this star topology, every link leads to exactly one receiver. At the end of the algorithm, the set of downstream receivers $\{R_i, 1 \leq i \leq N\}$ with isolated rates $\{r_i, 1 \leq i \leq N\}$ will be split into K subgroups $G_0, G_1, \ldots, G_{K-1}$ with reception rates $g_0, g_1, \ldots, g_{K-1}$. We put $\rho = (a+1)/(b+1)$ where a and b are parameters of algorithm 1.

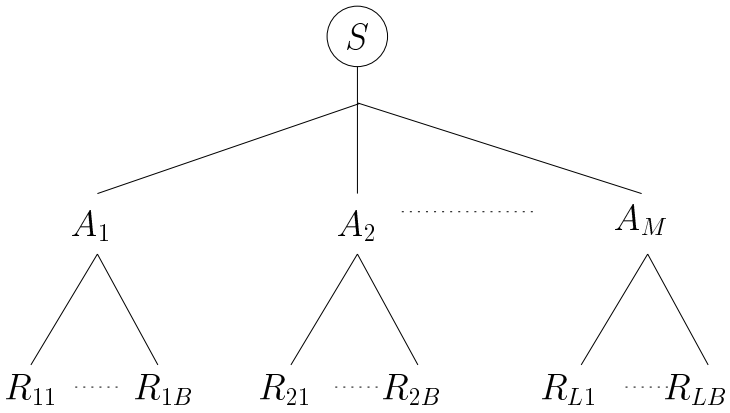


FIG. 6 – Network Model.

3.1.1 Algorithm Properties

Definition 1 (consecutive receivers) *We consider two receivers R_i and R_j as consecutive (in this order) if their corresponding isolated rates r_i and r_j satisfy $r_i < r_j$ and $\forall k$ ($k \neq i$ and $k \neq j$) : $r_k < r_i < r_j$ or $r_i < r_j < r_k$.*

Lemma 1 *If $r_j/r_{j+1} < \rho$, then when the algorithm converges, the consecutive receivers R_j and R_{j+1} will not belong to the same subgroup.*

Proof: Let γ be the multiplicative factor by which the source rate is multiplied at every period. Suppose that at a given time the two receivers R_j and R_{j+1}, belong to the same group, this means that none of their RTT variations is greater than the b threshold, that is : $\Delta\dot{\tau}_j = (r - r_j)/r_j < b$ and $\Delta\dot{\tau}_{j+1} = (r - r_{j+1})/r_{j+1} < b$. At the next period the source rate becomes γr rather than r. Receivers R_j and R_{j+1} will no longer continue to be in the same subgroup if $\Delta\dot{\tau_j}^+ = (\gamma r - r_j)/r_j > b$ and $\Delta\dot{\tau}^+_{j+1} = (\gamma r - r_{j+1})/r_{j+1} < a$, where $\Delta\dot{\tau_j}^+$ and $\Delta\dot{\tau}^+_{j+1}$ are their new relative RTT variations. It follows that $(b+1)r_j < \gamma r < (a+1)r_{j+1}$. Hence $r_j/r_{j+1} < (a+1)/(b+1) = \rho$. ■

Lemma 1 gives the necessary condition for two consecutive receivers to do not belong to the same subgroup at the end of the partitioning algorithm. It follows the next corollary :

Corollary 1 *Two consecutive receivers R_j and R_{j+1} will belong to the same subgroup if, $r_j/r_{j+1} \geq \rho$ or equivalently, $(1 + \Delta\dot{\tau}_{j+1})/(1 + \Delta\dot{\tau}_j) \geq \rho$*

Proof: The proof is straightforward using lemma 1. Knowing that $r/r_j = 1 + \Delta\dot{\tau}_j$ and $r/r_{j+1} = 1 + \Delta\dot{\tau}_{j+1}$, then the sufficient condition can be written : $(1 + \Delta\dot{\tau}_{j+1})/(1 + \Delta\dot{\tau}_j) \geq \rho$ ■

Theorem 1 *(Convergence Criteria) The partitioning algorithm 1 converges if (i) G subgroups are built, (ii) $|P_0| = 1$ or (iii) $\forall j < N, R_j \in P_0, \frac{1+\Delta\dot{\tau}_{j+1}}{1+\Delta\dot{\tau}_j} \geq \rho$*

Proof: The first criterion is enforced by the algorithm inputs while the second is obvious. The third criterion is a generalization of corollary 1 for every two consecutive receivers in P_0. ■

Theorem 2 *(Lower bound guarantee on the utility function) In a fully reliable multicast session, the overall utility that results from the execution of algorithm 1 has a lower bound expressed as : $\sum_{k=0}^{K-1} \alpha_{ik} \frac{1-\rho^{n_k}}{1-\rho}$*

This theorem shows that depending on ρ, the execution of the partitioning algorithm guarantees a lower bound on the session utility function independently of the receivers' isolated rates distribution.

Proof: Assuming a fully reliable multicast, the reception rate of the G_k's receivers is $g_k = \min_{R_i \in G_k} r_i = r_{k1}$. The G_k subgroup utility $U_k(g_k) = U_k(r_{k1}) = \alpha_{ik} \sum_{i=1}^{n_k} r_{k1}/r_{ki}$. For every two receivers R_i and R_j of subgroup G_k, we have $r_{i+j}/r_i = r_{i+j}/r_{i+j-1} \times r_{i+j-1}/r_{i+j-2} \times \ldots \times r_{i+1}/r_i \geq \rho^j$ since for every two consecutive receivers R_i and R_{i+1}, we have $r_{i+1}/r_i \geq \rho$. It follows that :

$$U_k(r_{k1}) \geq \alpha_{ik}\left(1 + \rho + \rho^2 + \ldots \rho^{n_k - 1}\right) = \alpha_{ik} \frac{1 - \rho^{n_k}}{1 - \rho}$$

Finally, the overall session utility satisfies :

$$U(g_0, g_1, \ldots, g_{K-1}) \geq \sum_{k=0}^{K-1} \alpha_{ik} \ \frac{1 - \rho^{n_k}}{1 - \rho}$$

■

3.1.2 Load Distribution

The main drawback of a replicated multi-rate scheme is that it wastes bandwidth due to data replication for every subgroup of receivers. We argue that our receiver-based approach allows for bandwidth saving compared to the classical sender-based scheme. In a sender-based replication, members of the same subgroup could be distributed among distant subtrees. This could result in data replication on multiple links of the multicast tree. We have shown in example 2 of section 2.3 how a receiver-based replication could be better in terms of consumed bandwidth in the presence of a hierarchy of routers. In the case of a star topology (which is one of the worst cases), we have the same consumed bandwidth. However, in terms of load distribution, a receiver-based replication allows for more load balanced bandwidth consumption among the different multicast tree links (computations are provided in appendix B). This has a great impact on the fairness properties of a receiver-based replication as will be studied in section 5.

Fig.7 plots the gain in load distribution when a receiver-based replication is adopted instead of a sender-based approach, as a function of the number of subgroups. The number of receivers per group is set to 1, 10 and 100 and rates are generated uniformally with parameters 5,55 and 25,35 giving the same mean but different mean variances and thus different heterogeneity degrees. First, we observe that the ratio is always greater than 1 justifying the benefit from the receiver-based replication on better load distribution. Moreover, the gain increases when increasing the number of subgroups. this is due to the fact that in a sender-based replication scheme, the source link will be more and more loaded since it transports all the data flows replicated by the source. Regarding the degree of heterogeneity, it is clear that the benefit from the receiver-based approach is more significant in the presence of more heterogeneous receivers.

3.2 Multiple Routers Topology

3.2.1 Lower Bound on the Utility Function

In a receiver-based replication, each router A_m $(1 \leq m \leq M)$ contributes to form $K-1$ subgroups $(Pkm, 1 \leq k \leq K-1)$ in addition to a subset of the source subgroup P_{0m}. Without loss of generality, assume that the built subgroups by a router are of the same size (B/K). Based on the minimum utility expression (theorem 2), the contribution of router A_m in the overall utility value satisfies[4] :

$$UR(g_1^m, g_2^m, \ldots, g_{K-1}^m) \geq \frac{K-1}{N} \ \frac{1 - \rho^{B/K}}{1 - \rho} \tag{7}$$

where $\forall i, k: \ \alpha_{ik} = 1/N$. The overall utility value when considering all the routers can then be bounded as follows :

$$UR(\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_{K-1}, g_0) \geq M \ \frac{K-1}{N} \ \frac{1 - \rho^{B/K}}{1 - \rho} + UR(g_0) \tag{8}$$

[4] We only consider the $K-1$ subgroups, since P_{0m} has to be considered once in the overall utility value.

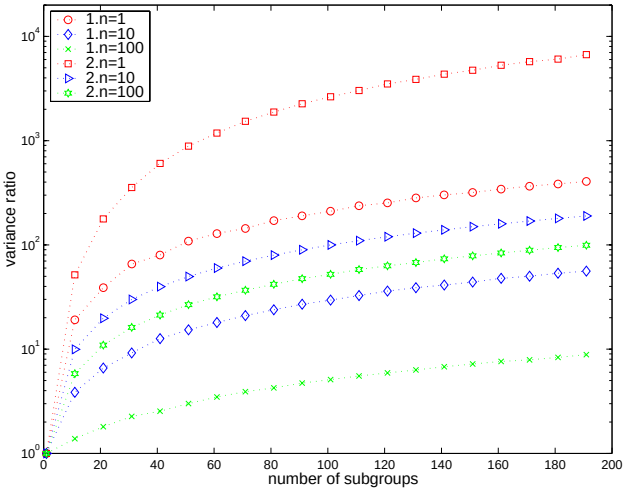


FIG. 7 – Gain in load distribution (1=[5,55] and 2=U[25,35]).

where $\underline{g}_k = (g_k^m)_{l=1..M,k=1..K-1}$ and g_k^m is the kth replication rate at router A_m. g_0 is the source transmission rate dictated by receivers of $P_0 = \bigcup_{l=1,M} P_{0m}$ and $UR(g_0)$ is the utility value contributed by the P_0 subgroup. Knowing that $UR(g_0) > 0$, we can write the following :

$$UR(\underline{g}_1^m, \underline{g}_2^m, \ldots, \underline{g}_{K-1}^m, g_0) > UR_{min} = M \ \frac{K-1}{N} \ \frac{1-\rho^{B/K}}{1-\rho} \tag{9}$$

3.2.2 Comparing with a source-based replication

The utility value in the sender-based replication scheme with K subgroups is (theorem 2) :

$$US(g_0, g_1, \ldots, g_{K-1}) \geq \frac{K}{N} \ \frac{1-\rho^{N/K}}{1-\rho} = US_{min} \tag{10}$$

The gain in utility in a receiver-based replication with respect to a source-based scheme can be expressed as follows :

$$\frac{UR_{min}}{US_{min}} = M \ \frac{K-1}{K} \ \frac{1-\rho^{B/K}}{1-\rho^{N/K}} \tag{11}$$

Assume that M and B are of fixed values, then UR_{min}/US_{min} decreases when K increases. For the largest value that K could take (B), we get :

$$lim_{B\to\infty} \ \frac{UR_{min}}{US_{min}} = lim_{B\to\infty} \ M \ \frac{B-1}{B} \ \frac{1-\rho}{1-\rho^M} = M \ \frac{1-\rho}{1-\rho^M}$$

$M \ \frac{1-\rho}{1-\rho^M}$ is an increasing function of M, that is, the gain of a receiver-based replication increases with the number of routers contributing in the execution of the partitioning algorithm. However, increasing the number of subgroups per router do not improve further the receivers satisfaction compared to a sender-based approach. Fig.8a plots for a fixed number of receivers, the gain as a function of K for different values of M. We easily observe that independently of the number of the routers, the gain in utility decreases when increasing the number of subgroups per router. Fig.8b plots the gain in utility when the number of receivers increases, with only 2 subgroups built per router. We can see that the gain in utility due to our receiver-based replication increases with the number of receivers. For instance with 2000 receivers located downstream 48 routers, we can improve the session utility with a factor of 20.

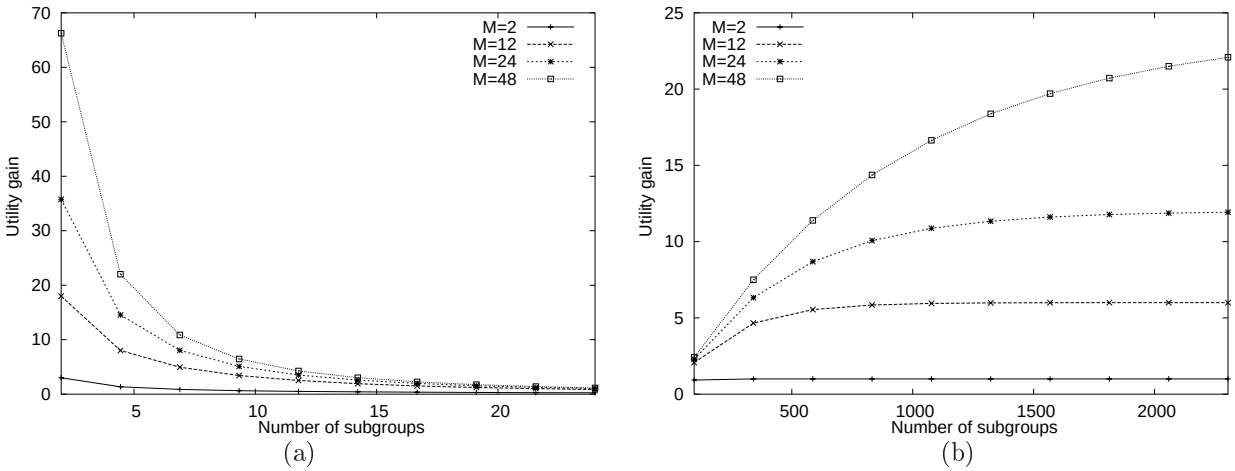


FIG. 8 – Utility gain of a receiver-based replication when increasing the number of (a) subgroups per router, $N = 2304$, (b) receivers, $K = 2$

4 Simulation Results

In this section, we evaluate our receiver-based replication scheme using simulations. For that, we use the same network model depicted in Fig.6. The receivers' isolated rates are generated following different distribution laws with different parameters, thus allowing for obtaining different heterogeneity degrees. The generated rates are randomly distributed among M sets of size B, each of which is associated to a router. The regulation tree construction algorithm is applied on these rates for different values of $\rho > 0.6$. Every simulation is repeated multiple times to allow for mean estimations of the different metrics of this study. First we are concerned with the benefit of the receiver-based approach compared to a single-rate approach. Afterwards, a comparison with the sender-based replication approach is provided.

4.1 Receiver-based Replication Versus Single-rate

In a single-rate congestion control, the source adapts its rate according to the worst receiver in the multicast group. Partitioning the receivers among subgroups allows for the source to increase its transmission rate. In our receiver-based replication scheme, the source transmits with a rate that matches the slowest receiver in P_0 (containing the fastest receivers) instead of the worst receiver in the whole multicast session. In order to show how the source transmission rate is never dictated by the worst receiver, we plot in figures 9a and 9b, the ratio of the sender rate in the receiver-based replication to the single rate scheme as a function of the number of receivers per router (B) and the number of routers (M) respectively.

We can see that independently of the number of routers and their receivers, the rate ratio is always greater than 1. For instance, Fig.9a shows that the source rate can be multiplied by 2.5 in the presence of 11 receivers per router with just a maximum number of subgroups $G = 3$ for a uniform distribution with parameter 5 and 55. Recall that G is the maximum number of subgroups per router which does not mean that every router effectively builds three subgroups. Moreover, we can see that the ratio is more significant when the heterogeneity degree is greater. For $G = 2$ and $B = 8$ (Fig.9a), the source rate is not improved significantly for the less heterogeneous set of rates (uniform parameters 5 and 10) while it is doubled for a more heterogeneous set of rates (uniform parameters 5 and 55). Fig.9a shows that the rate ratio decreases with the number of receivers per router. This is not a drawback of our approach since we estimate that the number of downstream links per router is limited. Furthermore, the ratio decreases also when the number of routers increases. This is quite normal, since increasing the number of routers increases the number of receivers that belong to the source subgroup P_0.

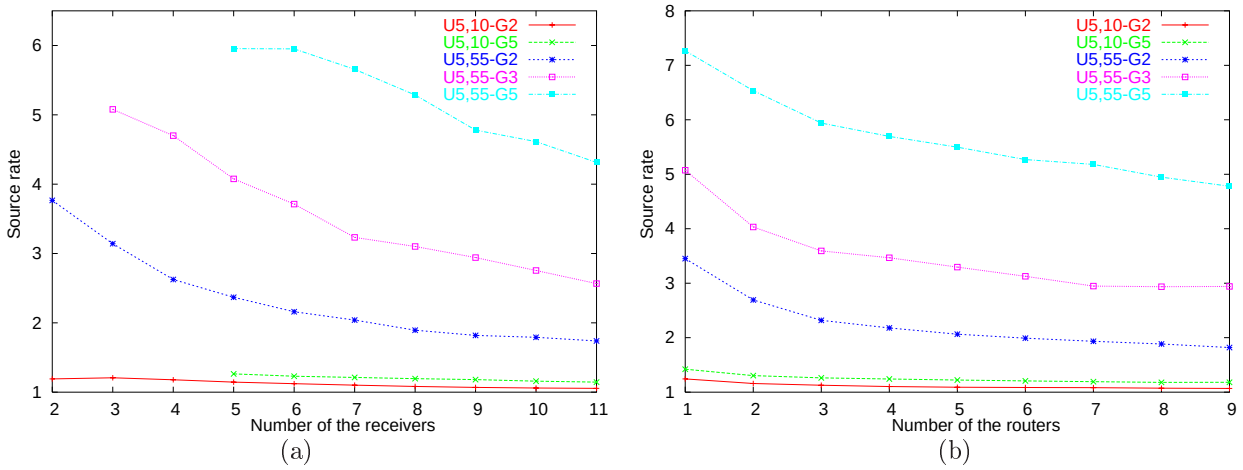


FIG. 9 – Source rate as a function of the number of (a) receivers per router, (b) routers.

4.2 Receiver-based Versus Sender-based Replication

Fig.10 plots the gain in utility due to the receiver-based approach as a function of the number of routers involved in the partitioning process for $B = 10$ and different G parameters. We can see that independently of the heterogeneity degree, the utility gain increases with the number of routers. For $G = 2$ and $M = 9$, the gain in utility is 1.4 when the receivers heterogeneity is greater (Fig.10b) instead of 1.04 (Fig.10a). Fig.10c plots the utility ratio for a set of rates distributed following an exponential law with mean 30 which is of a greater heterogeneity degree than the previous uniform distributions. Once again, we see that the gain is more significant. For instance for $G = 2$, we get a gain of 2.4 instead of 1.4 and 1.04 for the two previous cases. Finally, Fig.10 shows that increasing G beyond a given threshold does not increase significantly the gain in utility.

In order to see the impact of increasing B, the number of receivers associated to each router, Fig.11 plots the gain in utility as a function of B for uniform rate distribution with parameters (5,55) and an exponential distribution with mean 30. We can see that there is always a gain which in most cases, increases with the number of receivers per router. For instance, with $G = 3$ and 11 receivers per router, we obtain a gain of 1.7 and 2.6 for the uniform distribution and the exponential distributions respectively.

Now, we investigate the influence of the overall number of receivers. Fig.12 plots for different rate distributions, the utility gain as a function of the number of receivers subscribed to the multicast session. We can see that for the least heterogeneous set of receivers (Fig.12a), the gain does not exceed 1.12 and is almost constant even if the number of receivers increases. For more heterogeneous receivers (Fig.12b), we observe that the benefit increases with the number of receivers which allows for more scalability. Once again, we observe that increasing the number of subgroups per router does not necessarily allow for a significant benefit.

4.3 AMCA Extension

In this section, we provide some simulation results when applying our receiver-based replication on AMCA [14], a single-rate congestion avoidance algorithm. AMCA has been extended using ns-2.1b8 (network simulator [7]) to support more heterogeneous receivers. For more details about AMCA and its extension, the interested reader can refer to [14] and [12] respectively. For practical considerations and ease of implementation, we chose to limit the number of subgroups maintained by an active router to 2. In this way, routers are not overloaded by the management of multiple subgroups. However, since every router performs locally its own partitioning procedure, the overall number of subgroups seen by the source can be much higher. This choice is also motivated by our analytical and simulation results of the previous sections where it was observed that one does not need to have a large number of subgroups per router to significantly improve the

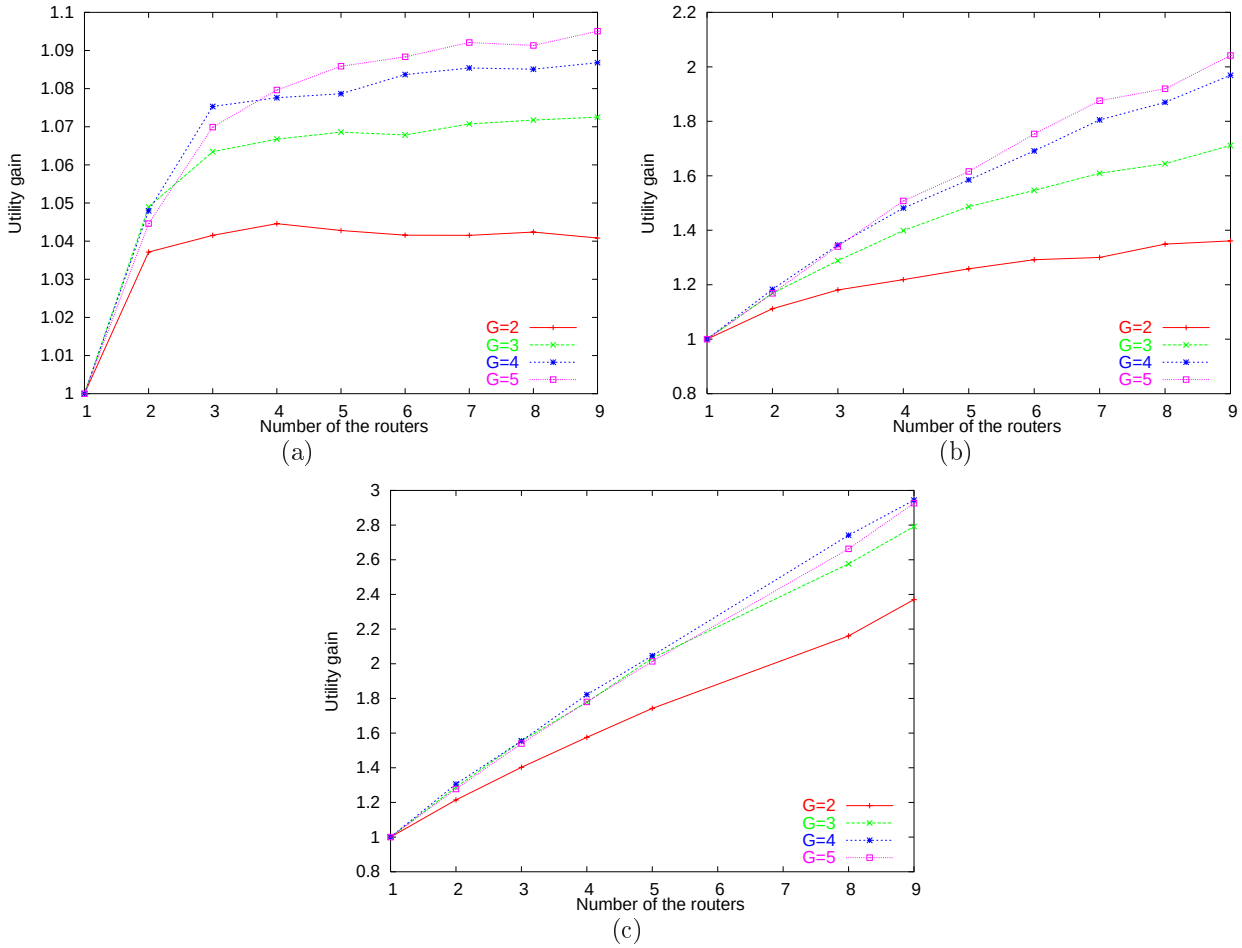


FIG. 10 – Utility gain as a function of the number of routers, $B = 10$, rates distributed following : (a) $U[5, 10]$, (b) $U[5, 55]$, and (c) E(30)

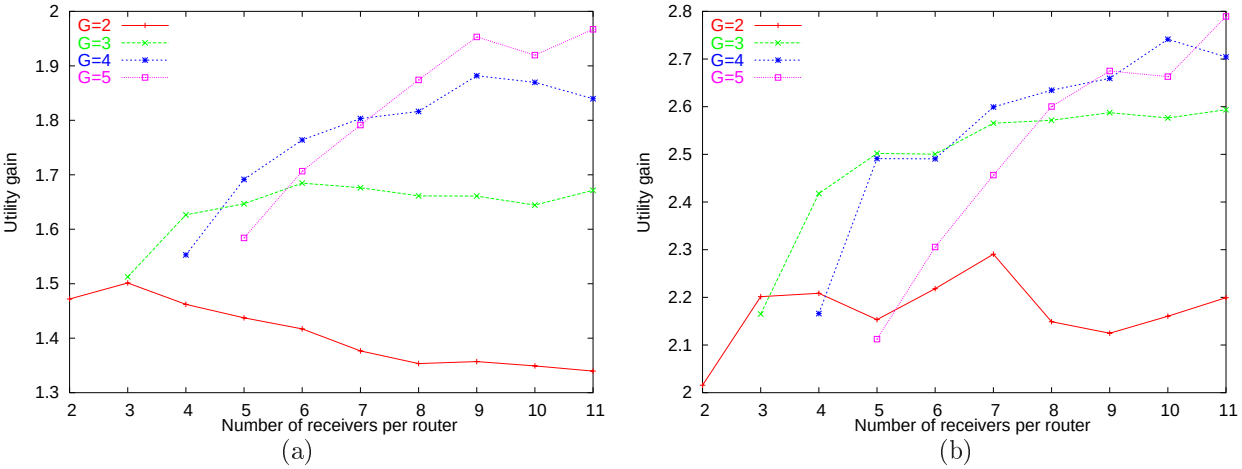


FIG. 11 – Utility gain as a function of the number of receivers per router (B), $M = 8$, (a) $U[5, 55]$, (b) $E(30)$

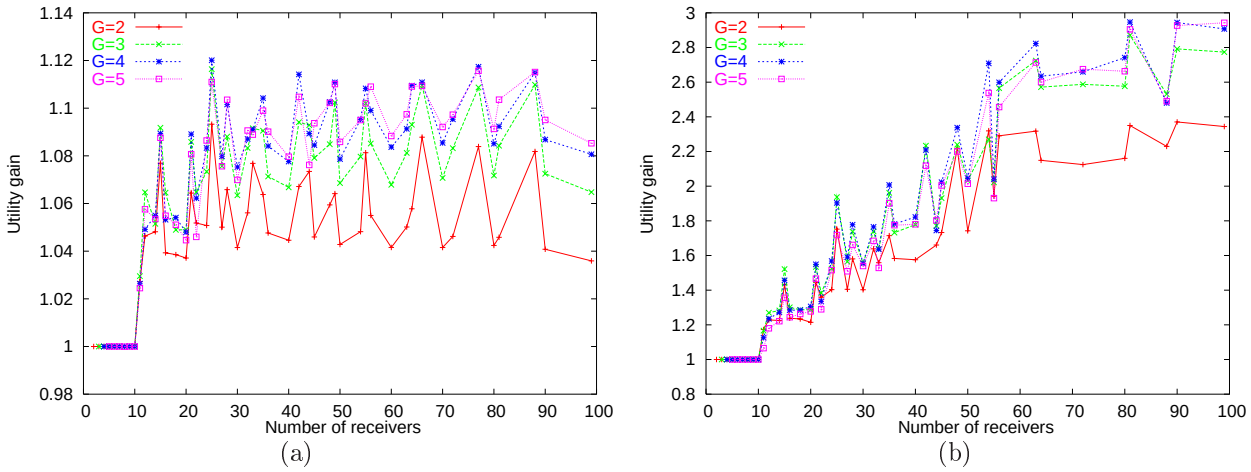


FIG. 12 – utility gain as a function of the number of receivers, (a) $U[5, 10]$, (b) $E(30)$

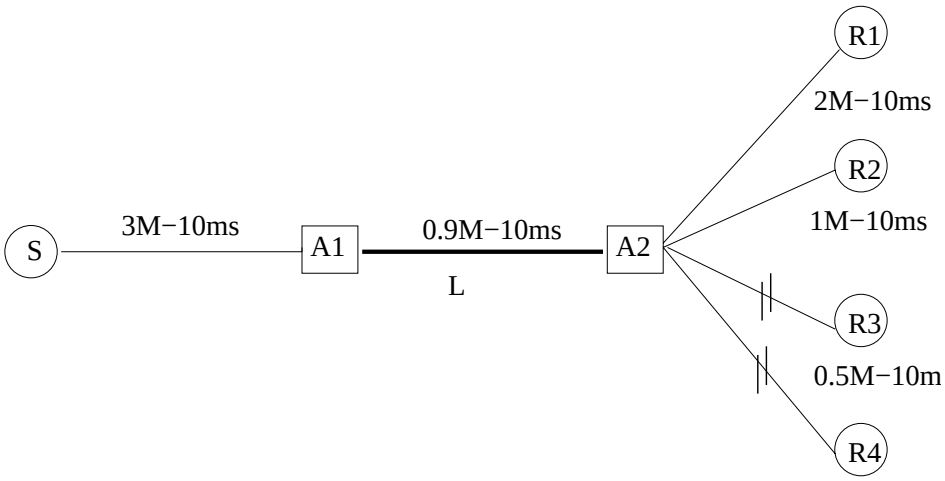


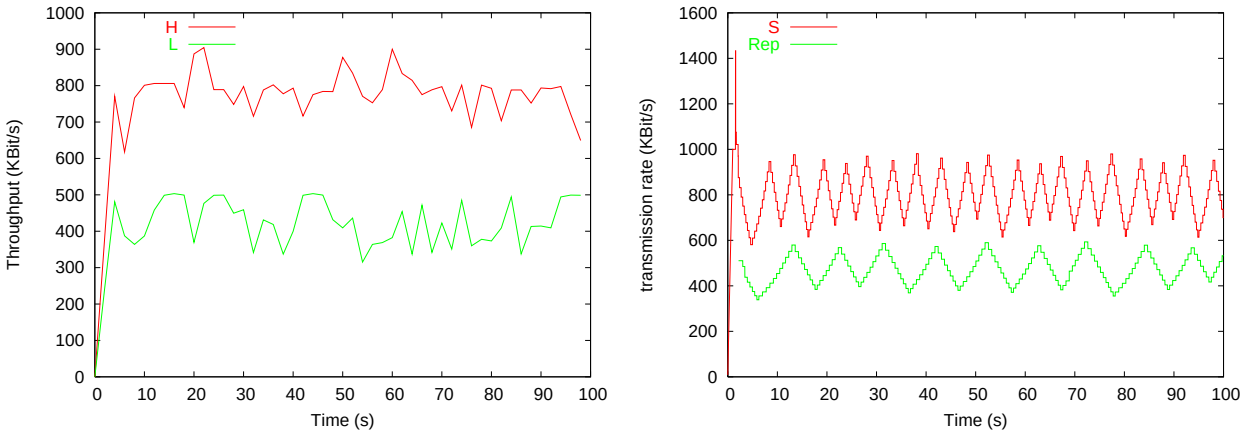
FIG. 13 – The used topology

receivers' satisfaction. It has also been proved formally in [8] that only 2 subgroups are sufficient to achieve a 50% of the best case (one receiver per subgroup), independently of the receivers heterogeneity.

A set of simulations have been conducted with a and b parameters set respectively to 0.05 and 0.2 ($\rho = 0.875$), on the network topology depicted in Fig.13. One source S multicasts data packets to 4 receivers (with isolated rate $0.9Mbps$ for $R1$ and $R2$, $0.5Mbps$ for $R3$ and $R4$) through two active routers A_1 and $A2$. The partitioning algorithm is enabled at the $A2$ router. Fig.14a shows the throughput achieved by the two subgroups $\{R1, R2\}$ and $\{R3, R4\}$ built by the active router $A2$. We can see that both the two subgroups' receivers obtain their isolated rates of $0.9Mbps$ and $0.5Mbps$. Fig.14b shows the transmission rates of the source and the chosen replicator (here $R1$). We can see that the source achieves rapidly a transmission rate of $900Kbps$, that the partitioning is performed in less than 2 seconds, and that the replicator achieves approximately a transmission rate of $500Kbps$ which corresponds to the isolated rate of the receivers of the lower capacity subgroup.

5 Inter-Session Fairness

In order to show how our replication scheme could enhance the inter-session as well as the intra-session fairness, we consider the star topology of Fig.15 where one multicast session shares one link (L) with K



(a) (b)

FIG. 14 – (a) Achieved throughput, (b) transmission rates

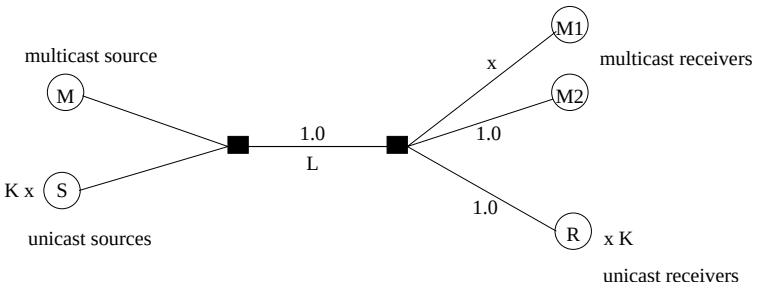


FIG. 15 – Fairness with unicast flows.

unicast sessions. The multicast source M multicasts data to 2 receivers M_1 and M_2. All the links have a bandwidth of 1.0 except the M_1 link which has a bandwidth of $x \leq 1.0$. We note by m_1, m_2 and s_i the (max-min) isolated rates of the receivers M_1, M_2 and R_i $(1 \leq i \leq K)$ respectively. Without loss of generality we put $\forall i, s = s_i$. In what follows, we consider three bandwidth allocation policies that result from applying a single-rate, a multi-rate source-based replication and a multi-rate receiver-based replication scheme. For a given policy, we note by $\widehat{m}_1$, $\widehat{m}_2$ and $\widehat{s}$, the allocated rates by this policy to the multicast and the unicast receivers.

To evaluate the inter-receiver fairness in the multicast session, we use as a fairness measure the utility function (1). For the considered topology of figure 15, the inter-receiver fairness of the multicast session can be computed using :

$$U = 0.5 \ (\frac{\widehat{m}_1}{m_1} + \frac{\widehat{m}_2}{m_2}) \tag{12}$$

where $\forall i$, α_i is set to 0.5. For the purpose of evaluating both unicast and multicast receivers utility (inter-session fairness), we use a global measure similar to the one proposed in [10] called the *global deviation measure*. The difference is that our metric measures the deviation of allocated bandwidth to the different receivers from their isolated rates, that is :

$$GD = \frac{\sum_{i=1}^{C_m} \sum_{k=1}^{n_i} \frac{r_{ki} - \widehat{r}_{ki}}{r_{ki}} + \sum_{j=1}^{C_u} \frac{s_j - \widehat{s}_j}{s_j}}{\sum_{i=1}^{C_m} (n_i)^y + C_u} \tag{13}$$

where C_u and C_m are the number of unicast and multicast sessions, n_i is the number of the receivers in the multicast session i, r_{ki} is the isolated rate of the kth receiver of the ith multicast session, s_j is the max-min rate of the jth unicast receiver, $\widehat{r}_{ki}$ and $\widehat{s}_j$ are the allocated rates for the kth receiver in the ith multicast

Policy	M_1	M_2	R	U	GD_0	GD_1
Isolated rates	0.1	0.25	0.3	-	-	-
Single-rate	0.1	0.1	0.3	0.7	0.15	0.12
Sender-based	0.1	0.2	0.23	0.9	0.1056	0.0844
Receiver-based	0.1	0.25	0.25	1.0	0.0417	0.0333

TAB. 2 – Star topology, $x = 0.1$ and $K = 3$

session and jth unicast session respectively and $y \in]0,1[$ is a parameter to obtain different relative weightings of multicast sessions.

5.1 Example

We consider a special case of the star topology (Fig.15) where the first multicast receiver link bandwidth x is set to 0.1 and the number of unicast sessions K is 3. This example is the same as the one given in [9]. Table 2 summarizes for each of the considered policies, the utility U achieved by the multicast session and the global deviation GD (GD_0 and GD_1 note the global deviation when y is set to 0 and 1 respectively).

In the presence of the three unicast receivers, M_1 would get an isolated rate of 0.1. M_2 would get $1/4 = 0.25$ if unconstrained by M_1. The three unicast receivers would get $(1 - 0.1)/3 = 0.3$. In a single-rate multicast, M_1 and M_2 gets a rate of 0.1 since they are limited by M_1. The three other receivers each get a rate of 0.3. In the case of a source-based replication scheme, the five receivers share the bottleneck bandwidth. That is, each of them would get 0.2. However M_1 is limited to 0.1 and thus the remaining bandwidth is supplied to the unicast receivers $(0.23 = (1 - 0.1 - 0.2)/3)$. In the receiver-based replication, M_1 is limited by 0.1 and M_2 competes fairly with the unicast receivers and all of them get a reception rate of 0.25. We observe that the multicast session utility achieves its highest value when the replication is performed by a receiver (here M_2). We also note that the minimum deviation is assured by our approach. All the receivers (unicast or multicast) get in general the reception rate which is the closest to their isolated rates. In what follows, we will use U_P and GD_P to note respectively the multicast session utility and the global deviation in the policy $P \in \{Sg, S, R\}$ for respectively a single-rate, source-based and receiver-based replication.

The isolated rates of the different receivers depend on the number of the unicast sessions K and the M_1's upstream link bandwidth x. The bottleneck link L is shared by $(K + 1)$ sessions (one multicast and K unicast sessions). Every session would get $1/(K + 1)$ of bandwidth unless $x < 1/(K + 1)$ (or $K < (1 - x)/x$). In this latter case, M_1 would get only x, M_2 will get $\frac{1}{K+1}$ and the unicast receivers will share the remaining bandwidth. As a result, we get :

$$m_1 = \begin{cases} x & \text{if } K < \frac{1-x}{x} \\ \frac{1}{K+1} & \text{otherwise} \end{cases} \qquad m_2 = \frac{1}{K+1}$$

$$s = \begin{cases} \frac{1-x}{K} & \text{if } K < \frac{1-x}{x} \\ \frac{1}{K+1} & \text{otherwise} \end{cases}$$

5.2 Single-rate Multicast

In a single-rate multicast, all the receivers subscribing to the same multicast session, get the same reception rate dictated by the worst receiver. Since the bottleneck link is shared with K other unicast sessions, we would get a reception rate of $1/(K + 1)$ for all of the receivers. However if $x < 1/(K + 1)$ then the multicast receivers would get a reception rate of x and the other unicast receivers would share the remaining bandwidth. We get the following :

$$\widehat{m} = \widehat{m}_1 = \widehat{m}_2 = \begin{cases} x & \text{if } K < \frac{1-x}{x} \\ \frac{1}{K+1} & \text{otherwise} \end{cases}$$

$$\widehat{s} = \begin{cases} \frac{1-x}{K} & \text{if } K < \frac{1-x}{x} \\ \frac{1}{K+1} & \text{otherwise} \end{cases}$$

giving,

$$U_{Sg} = \begin{cases} \frac{xK+(x+1)}{2} & \text{if } K < \frac{1-x}{x} \\ 1 & \text{otherwise} \end{cases} \tag{14}$$

and :

$$GD_{Sg} = \begin{cases} \frac{1-xK-x}{2^y+K} & \text{if } K < \frac{1-x}{x} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

5.3 Multi-rate Source-based Replication

In a source-based replication scheme, the bottleneck bandwidth would be fairly shared by the $(K+2)$ receivers. Thus, each of them would get a reception rate of $1/(K+2)$. However if $m_1 = x < 1/(K+2)$ or $K < (1-2x)/x$, then M_1 would be limited to x and the other unicast receivers would get back the remaining bandwidth $(1-1/(K+2)-x)/K$. In the opposite case ($K \geq (1-2x)/x$), the multicast receivers would get the same rate $(1/(K+1))$ as the other unicast receivers. In this case there is no replication from the source since M_1 and M_2 would have the same rate. In summary :

$$\widehat{m}_1 = \begin{cases} x & \text{if } K < \frac{1-2x}{x} \\ \frac{1}{K+1} & \text{otherwise} \end{cases} \qquad \widehat{m}_2 = \begin{cases} \frac{1}{K+2} & \text{if } K < \frac{1-2x}{x} \\ \frac{1}{K+1} & \text{otherwise} \end{cases}$$

$$\widehat{s} = \begin{cases} \frac{1-x}{K} - \frac{1}{K(K+2)} & \text{if } K < \frac{1-2x}{x} \\ \frac{1}{K+1} & \text{otherwise} \end{cases}$$

giving the following utility for the multicast session :

$$U_S = \begin{cases} \frac{2K+3}{2(K+2)} & \text{if } K < \frac{1-2x}{x} \\ 1 & \text{otherwise} \end{cases} \tag{16}$$

The global deviation can be expressed as follows :

$$GD_S = \begin{cases} \frac{2-x}{(2^y+K)(1-x)(K+2)} & \text{if } K < \frac{1-2x}{x} \\ \frac{x}{(2^y+K)(1-x)} & \text{if } K = \frac{1-2x}{x} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

5.4 Multi-rate Receiver-based Replication

In our receiver-based replication scheme, unicast receivers and the best one (M_2) in the multicast session would share the available bandwidth. M_1 will be regulated by M_2 and would get a reception rate of x. This holds when the reception rates of M_1 and M_2 are not the same. When $m_1 = x > 1/(K+1)$ (or $K < (1-x)/x$), then M_1 and M_2 would be in the same group and each of them would get $1/(K+1)$. Finally :

$$\forall K,\ \widehat{s} = \widehat{m}_2 = \frac{1}{K+1}$$

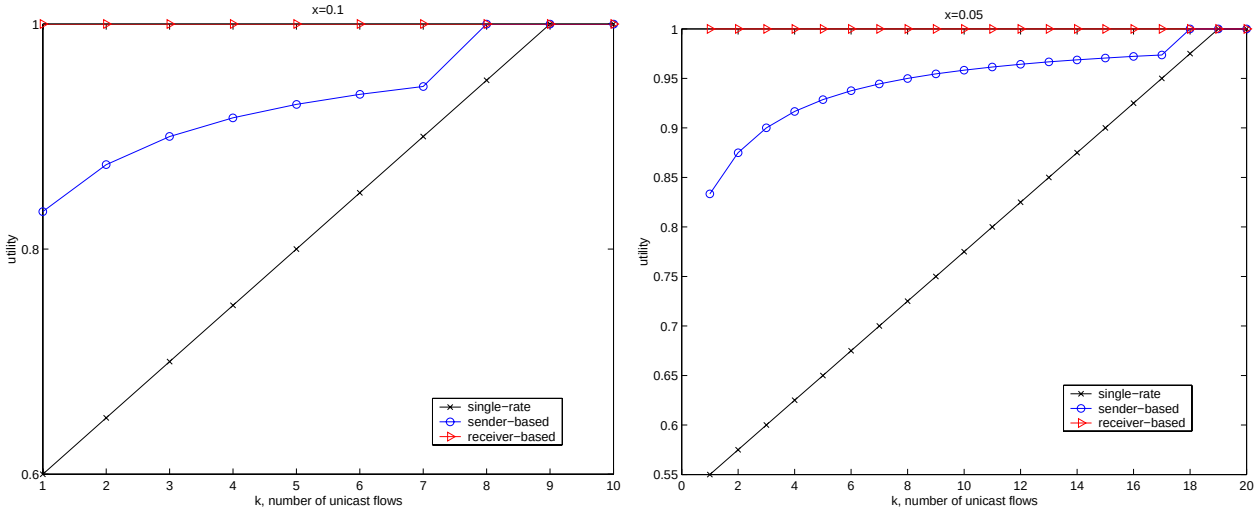
$$\widehat{m}_1 = min(\widehat{m}_2, x) = \begin{cases} x & \text{if } K < \frac{1-x}{x} \\ \frac{1}{K+1} & \text{otherwise} \end{cases}$$

giving a utility value of :

$$U_R = 1, \forall x, K \tag{18}$$

The global deviation can be expressed as follows :

$$GD_R = \begin{cases} \frac{1-(K+1)x}{(K+1)(1-x)(2^y+K)} & \text{if } K < \frac{1-x}{x} \\ 0 & \text{otherwise} \end{cases} \tag{19}$$



(a) (b)

FIG. 16 – Multicast session utility (a) $x = 0.1$ and (b) $x = 0.05$.

5.5 Numerical Results

We first examine the impact of the different approaches on the inter-receiver fairness in the multicast session. We plot in Fig.16, the utility value achieved by the multicast session as a function of the number of the competing unicast sessions for $x = 0.1$ (Fig.16a) and $x = 0.05$ (Fig.16b). We can see that the receiver-based replication scheme achieves its maximum value independently of K and x. This is due to the fact that we have only two receivers and a 2-subgroup partitioning is sufficient to get this maximum utility. Even with only two receivers, the source-based replication scheme achieves the maximum utility only in the presence of a minimum number of unicast flows ($K = 8$ when $x = 0.1$ and $K = 18$ when $x = 0.05$). These values of K corresponds to the minimum number of unicast sessions from which there is no data replication, since the allocated rates to the multicast receivers is the same ($K \geq (1 - 2x)/x$). We can also see that the single-rate approach is the worst one. As we will see, unicast sessions could be very aggressive toward single-rate multicast sessions.

Fig.17 plots the global deviation for each of the three policies. x is set to 0.1 in Fig.17a and 0.05 in Fig.17b while y is set to 0.5. We observe that independently of the number of unicast flows (K) and the isolated rate (x) of the slowest receiver in the multicast session, our approach incurs the smallest deviation from the isolated rates of the different receivers. This shows that our approach allows for more inter-session fairness than the two other approaches. The largest deviation is introduced by the single-rate scheme where the unicast sessions would be aggressive with the multicast session. The global deviation decreases with the number of the unicast flows since each unicast receiver would get a smaller rate, but closer to its isolated rate.

6 Conclusion

In order to accommodate heterogeneity in a multicast session, we proposed a new replication mechanism to implement a fine-grained multi-rate congestion control. Our approach consists in implying a set of receivers to replicate data they receive to other receivers with lower capacities. A replicator in the same way as a single-rate source, will adapt its rate depending on feedback it receives from the members of its associated subgroup. A partitioning algorithm is provided to split a set of receivers into subgroups of similar capacities. The main feature of this algorithm in addition to its simplicity, is that it does not rely on a prior knowledge of the receivers' capacities. The partitioning is performed on-the-fly as soon as feedback from the receivers are collected. The knowledge of the RTT variation experienced by every receiver is required but there is no assumption on how the RTT variations are measured, therefore a simple ping method could be suitable.

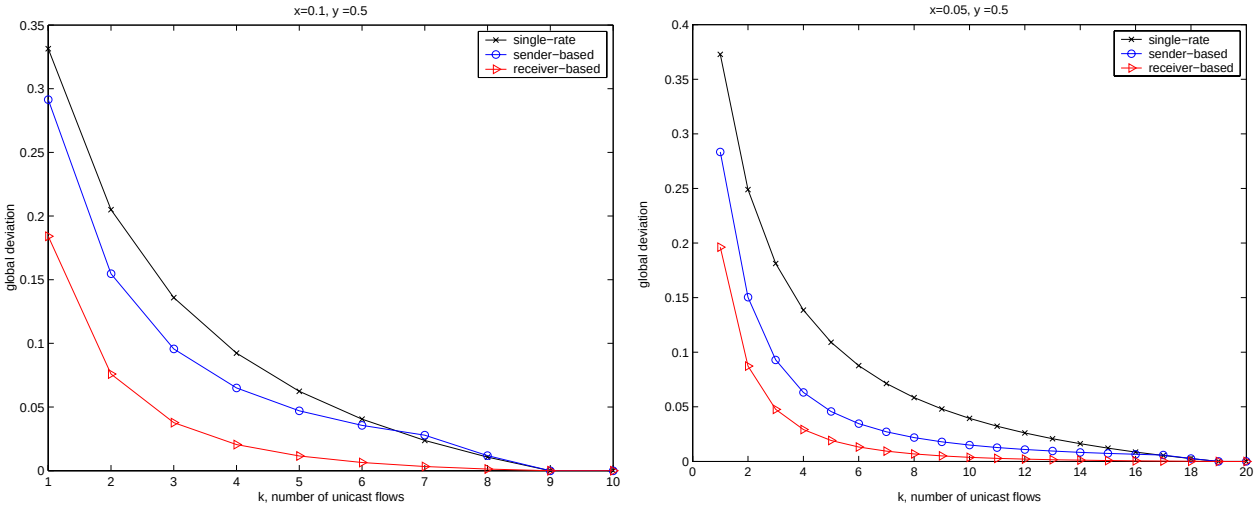


FIG. 17 – Global deviation, $y = 0.5$ and (a) $x = 0.1$, (b) $x = 0.05$.

Moreover, it guarantees a minimum utility value (depending on the ρ parameter), and with a minimum computation effort, it converges, or at least approximates the optimal solution.

To be more scalable and fairer with other sessions while improving the receivers satisfaction, we suggest to execute the partitioning algorithm at the routers. This additionally, allows for the construction of a regulation tree close to the multicast delivery one. Analysis and simulations showed that, our approach compared to single-rate and source-based replication schemes, is more scalable and allows for a better intra-session fairness. Compared to a layered approach, ours allows for a fine-grained congestion control without requiring a large number of subgroups. Using ns to validate our approach, an extension of a single-rate congestion avoidance algorithm (AMCA) has been implemented. Preliminary simulation results showed that the partitioning algorithm converges rapidly. As a future work, we plan to perform other simulations with more complex topologies, mainly to evaluate the dynamic behavior of our approach in the case of receivers changing their capacities over the time.

Références

[1] M. D. Amorim, O. C. M. B. Duarte, and G. Pujolle. Improving user satisfaction in adaptive multicast video. *IEEE/KICS Journal on Communications and Networks*, 4(3), September 2002.

[2] D. B. Bertsekas and R. Gallagher. *Data Networks*. Prentice Hall, 1992. Second Edition.

[3] Supratik Bhattacharyya, Donald F. Towsley, and James F. Kurose. The loss path multiplicity problem in multicast congestion control. In *INFOCOM'99*, pages 856–863, 1999.

[4] Nicolas Bonmariage and Guy Leduc. Adaptation dynamique des debits des couches pour la transmission video multipoint. In *Actes du Colloque Francophone sur l'Ingenierie des Protocoles (CFIP'2002*, 2002.

[5] J. Byers, M. Frumin, G. Horn, M. Luby, M. Mitzenmacherand A. Roetter, and W. Shaver. Flid-dl : Congestion control for layered multicast. In *NGC*, pages 71–81, November 2000.

[6] Dah Ming Chiu, Stephen Hurst, Miriam Kadansky, and Joseph Wesley. Tram : A tree-based reliable multicast protocol. Technical Report TR-98-66, SUN, July 1998.

[7] K. Fall, K. varadhan, and S. floyd. Ns notes and documentation ucb/lbnl/vint. software and documentation available at http ://www.isi.edi/nsnam/ns, July 1999.

[8] Rung-Hung Gau, Zygmunt J. Haas, and Bhaskar Krishnamachari. On multicast flow control for heterogeneous receivers. *IEEE/ACM Transactions on Networking*, 10(1) :86–101, February 2002.

[9] T. Jiang, M. Ammar, and E. Zegura. Inter-receiver fairness : A novel performance measure for multicast ABR sessions. In *Measurement and Modeling of Computer Systems*, pages 202–211, June 1998.

[10] T. Jiang, M. Ammar, and E. Zegura. On the use of destination set grouping to improve inter-receiver fairness for multicast abr sessions. In *IEEE INFOCOM'00*, March 2000.

[11] T. Jiang, E. Zegura, and M. Ammar. Inter-receiver fair multicast communication over the internet. In *the 9th International Workshopon Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, pages 103–114, June 1999.

[12] M. Maimour and C. Pham. Dealing with heterogeneity in a fully reliable multicast protocol. In *the 11th IEEE International Conference on Networks (ICON 2003)*, Sydney, Australia, September 28th - October 1st 2003.

[13] M. Maimour and C. Pham. A rtt-based partitioning algorithm for a multi-rate reliable multicast protocol. In *the 6th IEEE International Conference on High Speed Networks and Multimedia Communications (HSNMC 2003)*, Estoril, Portugal, July 23-25 2003.

[14] Moufida Maimour and Cong-Duc Pham. Amca : an active-based multicast congestion avoidance algorithm. In *the 8th IEEE Symposium on Computers and Communications (ISCC 2003)*, pages 747–754, Kemer-Antalya, Turkey, July 2003.

[15] T. Montgomery. A loss tolerant rate controller for reliable multicast. Technical Report NASA-IVV-97-011, NASA/WVU, August 1997.

[16] L. Rizzo. pgmcc : a tcp-friendly single-rate multicast. In *SIGCOMM*, pages 17–28, 2000.

[17] Dan Rubenstein, Jim Kurose, and Don Towsley. The impact of multicast layering on network fairness. *IEEE/ACM Transactions on Networking*, 10(2), April 2002.

[18] S. Shenker. Making gred work in networks : A game-theoretic analysis of switch service disciplines. In *SIGCOMM*, pages 47–57, 1994.

[19] H. Tzeng and K. Siu. On max-min fairness congestion control for multicast abr service in atm. *JSAC*, 15(3), April 1997.

[20] L. Vicisano, L. Rizzo, and J. Crowcroft. Tcp-like congestion control for layered multicast data transfer. In *Conference on Computer Communications (IEEE Infocom'98), San Fransisco, USA*, pages 1–8, 1998.

[21] Huayan Amy Wang and Mischa Schwartz. Achieving bounded fairness for multicast and TCP traffic in the internet. In *SIGCOMM*, pages 81–92, 1998.

[22] Y. Yang, M. Kim, and S. Lam. Optimal partitioning of multicast receivers. In *Proceedings of the 8th IEEE International Conference on Network Protocols (ICNP)*, Osaka, Japan, November 2000.

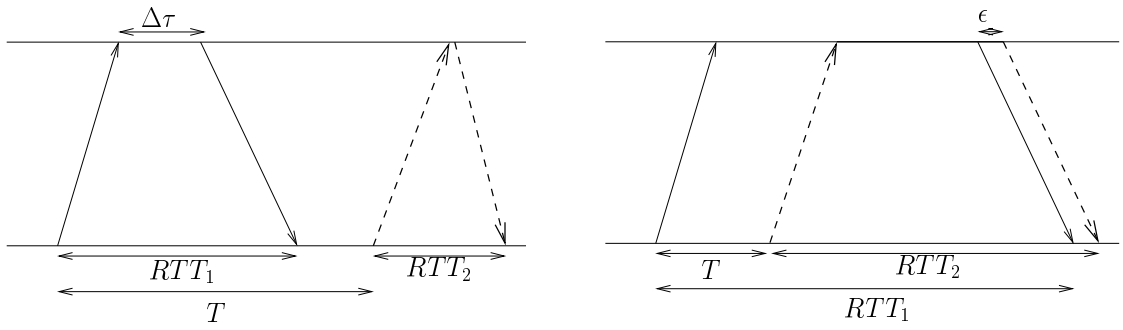
Appendices

A Utility Function Definition

Assume that $\Delta\dot{\tau}_i < 0$ for a receiver R_i, then $r < r_i$). We have to show that $\Delta\dot{\tau}_i$ could never be less than -1. Two cases are possible, either the next probing packet is sent after (Fig.Aa) or before (Fig.Ab) the reception of the previous probing packet. Remember that T is the probing period and let RTT_1 and RTT_2 be the last computed RTTs computed at the reception of the 2 last probing packets. In the first case (Fig.Aa), $|\Delta\tau|$ could never be greater than the probing period T. That is $T > |\Delta\tau|$ then $\Delta\dot{\tau}_i > -1$. In the second case (Fig.Ab), the second probing packet can not arrive before the first one, that is $\exists\epsilon > 0, |\Delta\tau| = RTT_2 - RTT_1 = T - \epsilon \Longrightarrow |\Delta\tau| < T$. Since $\Delta\dot{\tau}_i < 0$, we certainly have $\Delta\dot{\tau}_i > -1$.

B Load Distribution

We consider the star topology of figure 19 with one source multicasts data through one router, to Kn receivers distributed on K equisized subgroups $G_1, G_2, \ldots, G_K$ with replication rates of $g_1 < g_2 < \ldots < g_K$.



(a) $T > RTT_1$ (b) $T < RTT_1$

FIG. 18 – Lower bound for $\Delta\tau$.

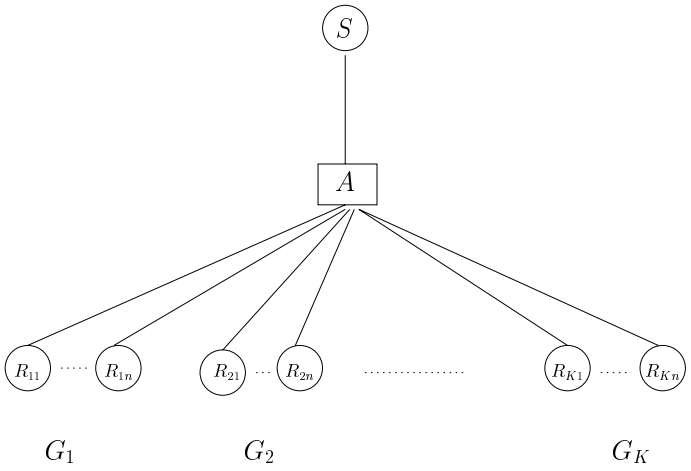


FIG. 19 – Star topology

X_S and X_R are two random variables that give the bandwidth consumption per link in a sender and a receiver-based replication scheme respectively. We derive the mean of the two random variables and their variation. This latter permits the evaluation of the load distribution among links in the two different approaches.

In a sender-based scheme, the consumed bandwidth on the source link equals the sum of all the replication rates $\sum_{i=1}^{K} g_i$. Every tail link of subgroup G_i is crossed by a flow of rate g_i. Noting by $L = N + 1 = nK + 1$, the total number of links, the mean consumed bandwidth can be expressed as follows :

$$\begin{aligned} E[X_S] = & \frac{(\sum_{i=1}^{K} g_i + n \sum_{i=1}^{K} g_i)}{L} \\ = & \frac{n+1}{L} \sum_{i=1}^{K} g_i \end{aligned} \tag{20}$$

In the case of a receiver-based replication, the consumed bandwidth on the source equals g_K, the highest replication rate. A replicator Rep_i $(i = 1..K-1)$ consumes its replication rate g_i in addition to its own reception rate g_{i+1}. The other receivers consume just their respective reception rate. All the n receivers of G_1 (the worst subgroup) receives data with rate g_1. In a subgroup G_i, $i = 2..K$, there is $n-1$ receivers each of them consumes g_i of bandwidth. Following that, we find that the X_R mean value is the same as for X_S :

$$E[X_R] = \frac{n+1}{L} \sum_{i=1}^{K} g_i \tag{21}$$

For the mean variation computation, we consider the case where we have at least two subgroups $(K > 1)$[5]. We have for X_S :

$$E[X_S^2] = \frac{(\sum_{i=1}^{K} g_i)^2 + n \sum_{i=1}^{K} g_i^2}{L} \tag{22}$$

thus, the mean variation can be computed as follows :

$$\begin{aligned} V[X_S] = & E[X_S^2] - E^2[X_S] \\ = & \frac{(\sum_{i=1}^{K} g_i)^2 + n \sum_{i=1}^{K} g_i^2}{L} - \frac{(n+1)^2}{L^2} (\sum_{i=1}^{K} g_i)^2 \\ = & \frac{(L - (1+n)^2)(\sum_{i=1}^{K} g_i)^2 + nL \sum_{i=1}^{K} g_i^2}{L^2} \end{aligned} \tag{23}$$

In the case of a receiver-based replication, we have :

$$E[X_R^2] = \frac{(n+1) \sum_{i=1}^{K} g_i^2 + 2 \sum_{i=1}^{K-1} (g_i g_{i+1})}{L} \tag{24}$$

and thus :

$$\begin{aligned} V[X_R] = & E[X_R^2] - E^2[X_R] \\ = & \frac{(n+1)L \sum_{i=1}^{K} g_i^2 + 2L \sum_{i=1}^{K-1} (g_i g_{i+1}) - (n+1)^2 (\sum_{i=1}^{K} g_i)^2}{L^2} \end{aligned} \tag{25}$$

[5] For $K = 1$, we have $V[X_R] = V[X_S] = 0$