A QoS-Aware Switching Mechanism Between The Two Modes of PIM-SM Multicast Routing Protocol

Fethi Filali and Walid Dabbous
INRIA Sophia-Antipolis, PLANETE Research Team
2004 Route des Lucioles, BP-93, 06902 Sophia-Antipolis, France
phone: (+33) 492387670, fax: (+33) 492387978
email:{filali,dabbous}@sophia.inria.fr

Abstract

PIM-SM is the only deployed intra-domain multicast routing protocol that builds both shared and source-based trees. However, it does not provide an efficient mechanism to switch between the two modes.

The PIM-SM switching mechanism proposed in this paper aims to improve both network and receivers requirements. It is a coordination-based mechanism in the sense that all concerned receivers contribute to the switching decision and not only the receiver requesting the switching. Furthermore, the mechanism uses information about the temporary available network resources provided by an underlying QoS-based unicast routing protocol or using periodic switching experiments to decide when to switch between the two modes.

Simulation results show that our mechanism achieves the inter-receiver switching fairness and that its integration in PIM-SM provides efficiently its original intentions.

Keywords: PIM-SM, QoS, switching mechanism, shared tree, source-based tree.

1 Introduction

PIM-SM is probably the most widely used multicast routing protocol today [2], [4]. PIM-SM creates a shared, RP-routed distribution tree that reaches all group members and it authorizes the receivers to switch from a RP (Rendez-vous Point)-routed tree (RPT) to a shortest path tree (SPT), however it does not specify how the switching policy should be done. The recommended policy in PIM-SM specification is to initiate the switch to the SP-tree after receiving a significant number of data packets during a specified time interval from a particular source. This heuristic for determining when to migrate from the RPT to the SPT, and vise versa, is far from being sufficient. There is no efficient provision in PIM for migrating from these two sub-modes. Individual routers make policy decisions as when to change the routing type for a given source.

To the best of our knowledge there is only one prior work describing an enhanced PIM-SM switching mechanism in the open literature [7]. Indeed, the authors have proposed an extension of PIM-SM called PIM-Switch which is based on the estimation of the density of the multicast group in the network. However, PIM-Switch has three main drawbacks. Firstly, it proposes the exclusive use either of RPT or SPT for all receivers[1]. Secondly, it does not take into account neither the QoS requirements of receivers nor the network requirements. Finally, there is no coordination between receivers to decide when and how to switch between the two modes of PIM-SM.

We believe that the use of a mechanism based on the coordination between the concerned receivers to switch between the RPT and the SPT will be more efficient and can fulfill to PIM-SM original intentions and make it more efficient comparing to other routing protocols and especially CBT [1].

The design of an efficient switching mechanism involves three major parts. The first part is the development of a decision algorithm that allows receivers to decide when they should request the switching from the RPT to the SPT and vice versa. When such decision is done, the second part involves the acceptance of the switching request by routers belonging to the path toward the source. This acceptance should take into account not only the receiver QoS requirements but also some network parameters such as the traffic concentration and the network resource usage. The third part deals with the syntax and the mechanism specification for making the switching decision and its integration in PIM-SM protocol.

In this paper, we explore the third parts. Clearly, we would like to improve the performance parameters that are of particular importance to the RP/network and to the receivers. To this end, we study and compare through simulation the tradeoff between the complexity and the effectiveness of the proposed switching mechanism.

The remainder of this paper is structured as follows. We start by giving the background and related work in Section 2. Some preliminaries, assumptions, and terminology will be presented in Section 3. In Section 4, we

[1] Throughout this paper we mean by a receiver, the designated router of at least one host connected to one of its directly-attached LANs and which is a member of a multicast group. We use the terms “receiver” and “designatd router”, interchangeably.

describe network and receivers-related parameters that should be taken into account during the design of our mechanism. Section 5 details our switching mechanism and new PIM messages used to handle the switching between the RPT and the SPT. The performance evaluation of the mechanism will be the subject of Section 6. Finally, Section 7 concludes our work and outlines some future issues.

2 Background and Motivation

2.1 Background

Similar to the CBT protocol [1], PIM-Sparse Mode (PIM-SM) [4] is designated to restrict multicast traffic to only those routers interested in receiving it. PIM-SM constructs a multicast distribution tree around a router called a rendezvous point (RP). This rendezvous point plays the same role as the core in the CBT protocol; receivers "meet" new sources at this rendezvous point. However, PIM-SM is a more flexible protocol than CBT. While CBT with trees are always group-shared trees, with PIM-SM an individual receiver may choose to construct either a group-shared tree or a shortest-path tree.

The PIM-SM protocol initially constructs a group-shared tree to support a multicast group. The tree is formed by the senders and receivers both connecting to the rendezvous point, just as a shared tree is constructed around the core with the CBT protocol. After the tree is constructed, a receiver (actually the router closest to this receiver) can opt to change its connection to a particular source to a shortest-path tree. This is accomplished by having this router send a PIM join message to the source. Once the shortest path from source to receiver is created, the extraneous branches through the RP are pruned. This procedure is illustrated in Figure 1. Note that different types of trees can be selected for different sources within a single multicast group.

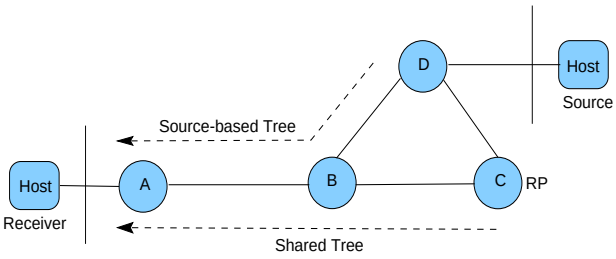

Figure 1: Example: Switching from shared tree to shortest path tree

There are advantages to each type of distribution tree. The shared tree is relatively easy to construct, and it reduces the amount of state information that must be stored in the routers. Accordingly, a shared tree would conserve network resources if the multicast group consisted of a large number of low-data-rate sources. However, as indicated above, shared trees cause a concentration of traffic around the core or the rendezvous point, a phenomenon that can result in performance degradation if there is a large volume of multicast traffic. Another disadvantage of shared trees is that traffic often does not traverse the shortest path from source to destination. If low latency is a critical application requirement, it would be preferable for traffic to be routed along a shortest path. PIM-SM architecture supports both types of distribution trees.

A PIM-SM router has at most three entry types in its multicast route table:

- A (*,*,RP) entry: a special entry type to support interoperability which must be supported by all PIM routers. A data packet will match on a (*,*,RP) entry if there is no more specific entry (such as (S,G) or (*,G)) and the destination group address in the packet maps to the RP listed in the (*,*,RP) entry. In this sense, a (*,*,RP) entry represents an aggregation of all the groups that hash to that RP.
- A (*,G) entry: a wildcard multicast route entry for the group.
- A (S,G) entry: a multicast route entry that is specific to the source.

The switching technique from the RPT to the SPT that is recommended in the PIM-SM specification [4] is a rate-based policy. The receiver initiates the switching to the SPT after receiving a significant number of data packets during a specified time interval from a particular source. When a (*,G), or corresponding (*,*,RP), entry is created, a data rate counter may be initiated at the last-hop routers. The counter is incremented with every data packet received for directly connected members of an SM group, if the longest match is (*,G) or (*,*,RP). If and when the data rate for the group exceeds a certain configured threshold t_1, the router initiates 'source-specific' data rate counters for the following data packets. Then, each counter for a source, is incremented when packets matching on (*,G), or (*,*,RP), are received from that source. If the data rate from the particular source exceeds a configured threshold t_2, a (S,G) entry is created and a Join/Prune message is sent towards the source. If the RPF interface for (S,G) is not the same as that for (*,G) - or (*,*,RP), then the SPT-bit is cleared in the (S,G) entry.

2.2 Motivation Example

A receiver's DR that wants to switch to the SPT should send a PIM (S,G) Join message toward the source. All receivers in the SP-subtree[2] will receive data from the SPT. The problem is that the QoS perceived by other receivers

[2]The SP-subtree is the part of the multicast tree from the SP, a specific on-tree router that will be defined in Section 3. It is the common part between the SPT and the RPT from the receiver initiating the switching, referred hereafter as the ISR (Initiator Switching Receiver).

in the SP-subtree can be affected when a receiver switches from the RPT to the SPT without coordinating with them.

To demonstrate the limitation of the PIM-recommended switching technique, let us consider the topology shown in Figure 2, where there is a multicast source S sending to four receivers R_1, R_2, R_3, and R_4 via the RP-routed tree (RPT). The capacity of the link between the designated router of receiver R_3 toward the RPT is equal to 36 kb/s, while that of receivers R_1, R_2, and R_4 is equal to 128 kb/s.

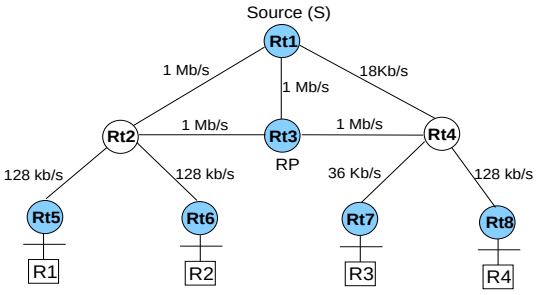

Figure 2: Motivation sample

We examine the impact of the PIM recommended switching mechanism on the application performance. We distinguish two cases: when the application uses a single-rate congestion control mechanism such as TFMCC protocol described in [11] and when it uses a layered transmission scheme such as WEBRC [10]. We assume that the designated router of receiver R_3 (router Rt_7) decides to switch from the RPT to the SPT because it receives a high data rate.

For single-rate multicast sessions, the source sends at the rate of the slowest receiver which is the receiver R_3 in our example. When switching to the SPT, the bottleneck in the tree will be the link between the source and router Rt_4 which has a capacity equal to 18 kb/s. As a result the sending rate will be reduced because of the higher loss rate in this bottleneck. If receiver R_3 knows in advance the resulting rate, it would be better to him to not switch to the SPT.

For multi-rate sources, we assume that the source S uses three layers, layer 1 with a rate equal to 9 kb/s sending to multicast group G_1, layer 2 with a rate equal to 18 kb/s sending to group G_2, and layer 3 with a rate equal to 36 kb/s sending to group G_3. We suppose that all group addresses are mapped to the same RP, router Rt_3 which is in an over-provisioned place. We assume that all receivers except R_3 are already subscribed to the three layers. Their aggregated receiving rate is then equal to 9+18+36 = 63 kb/s. Receiver R_3 can subscribe only to the two first layers and it switches to the SPT. As consequence, receiver R_4 will detect a high loss rate and it will be obliged to leave layer 2 and layer 3. This decreases considerably the quality of received data.

3 Preliminaries, Assumptions and Terminology

We represent a network by a weighted diagraph (V, E), where V denotes the set of network nodes and E the set of communication links that connect the nodes. $|V| = N$ and $|E| = M$ denote the number of nodes and links in the network, respectively. $P(u, v)$ represents the set of routers constituting the path between u and v (including u and v).

Associated with each link are parameters that describe the current status of the link, for example the average link delay, the link cost, the loss rate, and the bandwidth available on the link. We term these parameters *link state*. Similarly, associated with each node are parameters representing the current status of the node, for example routing tables. We term these parameters *node state*. The set of links states and nodes states constitutes the *network state*. We assume that all routers in the network support PIM-SM [4] multicast routing protocol.

During a multicast session, we assume that each receiver controls the quality of reception parameters and have a knowledge of the required values that would be accepted. For example, the receiver R should specify R_R, P_R, D_R referring to the minimum data rate, the maximum loss rate, and the maximum delay, respectively. These parameters values may differ from receiver to another.

In addition, each receiver maintains the values of QoS parameters of data received from the multicast tree, we denote by $R_R(RPT)$, $P_R(RPT)$, and $D_R(RPT)$ the current data rate, loss rate, and delay from the RPT, respectively. Depending on the nature of the multicast application, one or more QoS parameters may be considered as an important measure of the reception quality.

The receiver that decides to switch to the SPT, hereafter called the **INITIATOR-SWITCHING RECEIVER (ISR)**, has to send a switching request toward the source. In its request, the member may specify the set of QoS parameters violated and their current and required values.

We use the term **SWITCHING-POINT (SP)** of a receiver R to refer to the last router in the path between R and a source S which belongs both to the RPT and the SPT. This router could also be defined as the first router between the receiver and the source which uses different upstream interfaces to reach the RP and the source.

Therefore, for a session (S, G), there is an unique SP for each receiver. A receiver may have the same SP to two different sources and two receivers may have different switching point to the same source. The SP may be the source itself if the RP belongs to the shortest path between the source and the switching receiver.

For example, for the multicast group shown in Figure 2, router Rt_2 is the SP of receivers R_1 and R_2 and router Rt_4 is the SP of receivers R_3 and R_4.

4 Switching Parameters

As outlined in [4], the switching to the SPT can be initiated either by the RP or by designated routers. Different criteria could be applied to trigger the switching from the RPT to the SPT. These criteria may be the QoS requirements and/or the network parameters. In this section, we explore both types of parameters.

4.1 QoS Receivers Parameters

Certainly that the first requirement of receivers is the Quality of Service (QoS) of the data received from multicast sources such as delay, bandwidth, and loss rate.

4.1.1 Delay bound

Let $d(i,j)$ be the delay of the link between node i and node j. Since the delay is an additive parameter, the delay $d(u,v)$ between node u and node v is equal to the sum of individual link delays along the path $P(u,v)$ between u and v:

$$d(u,v) = \sum_{(i,j)\in P(u,v)} d(i,j) \tag{1}$$

Each receiver R can estimate and compare the mean delay of data received from the RPT to the required delay. The delay bound inequality is:

$$D_R(RPT) \leq D_R \tag{2}$$

where D_R is the minimum delay required by the receiver R and $D_R(RPT)$ is the mean delay from the RPT.

4.1.2 Bandwidth bound

Let $b(u,v)$ denote the available bandwidth along the path between node u and node v. The bandwidth is a concave parameter and then $b(u,v)$ is computed as follows:

$$b(u,v) = \min_{(i,j)\in P(u,v)} b(i,j) \tag{3}$$

where $b(i,j)$ is the available bandwidth in the link between node i and node j. The rate bound inequality is:

$$R_R(RPT) \geq R_R \tag{4}$$

where R_R is the minimum required rate by the receiver R and $R_R(RPT)$ is the actually available rate from the RPT.

4.1.3 Loss rate bound

We note by $L(i,j)$ the loss probability of the link between node i to node j. The loss rate is a multiplicative parameter, the probability $P(u,v)$ that a packet sent by router i arrives to node j is equal to the product of no loss probability of links between u and v.

$$P(u,v) = \prod_{(i,j)\in P(u,v)} (1 - L(i,j)) \tag{5}$$

The loss bound inequality is:

$$P_R(RPT) \geq P_R \tag{6}$$

where $(1 - P_R)$ is the maximum loss rate required by the receiver R and $(1 - P_R(RPT))$ is the actually loss rate from the RPT.

4.2 Network Parameters

The shared tree PIM-SM's mode is expected to concentrate traffic onto the subset of network links that compose the shared trees. In contrast, the source-based tree PIM-SM's mode is expected to distribute the traffic more evenly among all links because it uses a different tree for each sender and each group [4].

One of the most important interest in using PIM-SM is its ability, via the bootstrap mechanism, to distribute the multicast traffic load between RP candidates. In fact, when a traffic concentration around RP routers holds, it may affect the multicast data delivery quality.

The traffic load around the RP depends on the sending rate of the sources in the multicast group that use the RPT. We assume that for each multicast group, the RP maintains the data rate received from all sources and the rate from individual source. We note by $R^i_{RP}(S,G)$, the rate of data received from source S belonging to group G on the multicast-enabled interface i of the RP. The total data rate of all sources of group G on interface i is given by:

$$R^i_{RP}(*,G) = \sum_{S\in\{RPT\}_G}^{i} R^i_{RP}(S,G). \tag{7}$$

where $\{RPT\}_G$ and $\{SPT\}_G$ are the set of sources of group G that use the shared tree and the source-based tree, respectively. The total rate of the multicast traffic received by the RP on the interface i is computed as follows:

$$R^i_{RP}(*,*) = \sum_{G} R^i_{RP}(*,G) \tag{8}$$

We assume that for each interface i the RP maintains a multicast traffic maximum fraction X_i (a multicast traffic threshold) of the link capacity C_i that should not be exceeded. Links where the multicast rate exceeds X_iC_i are considered over limit.

5 Enhanced PIM-SM Switching Mechanism

We describe in this section a switching mechanism based on one or more parameters from those enumerated in the previous section. We first outline our mechanism

overview. Then, we detail it by exploring the new proposed PIM messages as well as some implementation issues.

5.1 Mechanism Overview

The simplest PIM-SM switching mechanism that we could use is to switch all receivers from the RPT to the SPT if at least one receiver decides to switch to. This is useful when we make the assumption that all the receivers receive data with the same quality of service. In this case each receiver will consider that the SPT is more efficient than the RPT. The advantage of such type of mechanism is that it is easy to implement, however it does not guarantee that all receivers even the ISR will be satisfied after the switching.

Another switching alternative is to consider only receivers belonging to the ISR's SP-subtree. In such mechanism, we assume that only SP-subtree receivers (and not all receivers) are concerned by the switching decision because is the SP which forwards to them the data received either from the RPT or from the SPT. The current version of PIM-SM uses implicitly this kind of mechanism [4].

We believe that both alternatives are not efficient because they do not take into account the receivers reception quality and their connectivity heterogeneity. Furthermore, there is no coordination between the receivers concerned by the switching.

Our mechanism is based on a coordination between the receivers concerned by the switching initiated by the ISR. Certainly, that such mechanism can be costly in terms of complexity and switching latency because it requires specific states to be maintained by at least the SP and designated routers but it has the advantage of taking into account the requirements of all concerned receivers by the switching decision. We will evaluate the tradeoff between both parameters in Section 6.

Each switching point should maintain a soft-state per multicast entry concerning previous received PIM switching requests. A timer is maintained for each state. After the expiration of the timer, the SP drops. The state is updated whenever it receives a new information concerning new PIM[3] Switch requests.

The receiver who wants to switch to the SPT should first verify if there is already a switching request under processing that have been sent by him or by another downstream receiver. If it is not the case, it sends immediately a Switch Request message toward the source. Otherwise, it should wait for the expiration of [Switch-Request-Send-Timer] before sending its request.

After receiving the Switch request, the SP router sends a coordination request to other interfaces in the *oif* list of the (*,G) entry. Receivers may accept or refuse the request. When the SP receives a Switch Accept message from all interfaces or the timer [Switch-Coordination-Wait-Timer] expires without receiving all responses, it sends a Switch Ack message to the ISR. If at least one of the receivers refuses the switching (sends a Switch Refuse message to the SP), the SP sends immediately a Switch Nack message to the ISR.

In the next section, we describe in detail our mechanism and the new proposed PIM messages. Details about the Switching Data Base and the timers default values can be found in [5].

5.2 Detailed Mechanism Description

Our mechanism can be included in PIM-SM using the following new messages:

- PIM Switch Request: this message is sent by the initiator switching receiver toward the source.
- PIM Switch Coordination: this message is sent by the Switching Point (the ISR's SP) to its downstream receivers to ask them to accept or to refuse the ISR's switching request.
- PIM Switch Accept/Refuse message: this message is sent to all SP's downstream receivers toward the SP.
- PIM Switch ACK message: this message is sent by the SP to downstream receivers in order to inform them that they are authorized to receive data directly from the source and then send a PIM Join (S,G) to the source.
- PIM-SM Switch NACK message: this message is sent by the ISR's SP in order to inform it that at least one of them did not accept to switch to the source.

We summarize in Table 1, the messages used by our switching mechanism.

Message	From –> To	When ?
Switch Request	ISR –> Source	The ISR wants to switch to SPT
Switch Coordination	SP –> Downstream Receivers	The SP receives a Switch Request
Switch Refuse	Receiver –> SP	The receiver refuses the switching
Switch Accept	Receiver –> SP	The receiver accepts the switching
Switch Ack	SP –> ISR	All downstream receivers accept the switching
Switch Nack	SP –> ISR	At least one downstream receiver refuses the switching

Table 1: Summary of the new PIM Messages

In the following sub-sections, we will give more details about these messages: when they are sent? What each router should do when receiving these messages? etc.

[3]Throughout this paper we use the term PIM to refer to the Sparse Mode (SM) of PIM (PIM-SM).

5.2.1 PIM Switch Request Message

A PIM Switch Request message is sent by the receiver that wants to switch to the SPT, so called the ISR (Initiator Switching Receiver), toward the source. This receiver should specify in its request message the QoS parameters needed and their current and requested values. After sending this request, the receiver continues receiving packets coming from the RP. It will send a PIM Join message to the source only after reception of the PIM Switch Ack message from its Switching Point (SP).

A PIM-SM router which receives a PIM Switch Request should first determine if he is the Switching Point (SP) of the receiver sending that message. If so, this router sends to the downstream receivers a PIM Switch Coordination message to inform them that there is one receiver who wants to switch to the SPT and that they may accept or refuse the switching.

Intermediate routers between the ISR and its SP have only to forward the packet toward the source.

When a router receives a PIM Switch request from a new ISR of which he is the SP, while he does not send a PIM switch Ack/Nack to the previous ISR, the router deletes this message.

To prevent from large amount of signaling switching messages, the receiver should send at most one PIM Switch Request each [Switch-Request-Time]. Each receiver maintains a history about switching requests that it has sent toward each active source.

When an ISR does not receive neither a Switch Ack nor a Switch Nack message, it assumes that its message was lost in the path towards the SP and it tries again after the expiration of the timer [Switch-Request-Time].

5.2.2 PIM Switch Coordination Message

When receiving a PIM Switch Request message, the SP should first verify whether it has local receivers or no. If it is the case it checks whether the switching to the SPT does not violate their QoS requirements. If so, this router sends a PIM-Switch Coordination message to its downstream interfaces belong to the *oif* list of the (*, G) entry except the interface from which it received the switching request. This message contains a copy of the requested parameters given by the ISR in its Switch request message.

A router that receives a PIM switch coordination message, should check whether it has local receivers or no. If so and when the eligibility tests[4] succeed, it forwards the message to other interfaces that belong to the *oif* list.

If the eligibility test do not succeed, the router should immediately send a PIM Switch Refuse message to the SP. In this case there is no need to forward the coordination message to other interfaces.

[4]The eligibility tests will be the subject of Section 5.4.

5.2.3 PIM Switch Accept/Refuse Message

When receiving a PIM Switch Coordination message, a receiver determines whether the switching to the source-based tree violates its reception quality or no. If so, it sends a PIM Switch Accept message to the ISR's Separation Point, otherwise it sends a PIM Switch Refuse message.

This message should be sent immediately to the separation point which has sent the PIM-SM coordination message. A router who receives a PIM switch Accept message should wait for the response from receivers downstream to other interfaces. It forwards this message only if it receives an accept message from every interface where there is at least one downstream receiver. In other cases, it sends a PIM Switch Refuse message.

The receiver which sends a PIM Switch Nack message may specify the reasons why it refuses the switching to the SPT.

A router which receives the PIM Switch Accept to forward to the ISR's Separation Point and which has already sent a PIM Switch Accept/Refuse of the same ISR before [Switch-Accept-Refuse-Timer] should reject this message. In contrast, when it receives a PIM Switch Refuse message, it should forward it to its destination.

The router can know if it has already sent a PIM Switch Accept/Refuse to the ISP by consulting its Switching Data Base where entries are kept alive for [Switch -SDB-entry-time].

5.2.4 PIM Switch Ack/Nack Message

After Sending a PIM Switch Coordination message, the SP should wait a maximum time equal to [Switch-Coordination-Timer] for downstream routers answers from all interfaces before sending the Switch Ack/Nack message to the ISR.

When the separation point receives a PIM Switch Accept from all interfaces in *oif* list except that from which it received the switching request, it sends a PIM Switch Ack towards the switching initiator receiver[5]. In the case when there is at least one PIM Switch Refuse message received from one of the downstream interfaces, the SP sends a PIM Switch Nack to the ISR without waiting other PIM switch Accept/Refuse from other interfaces.

For security reasons, upon receiving a PIM Switch Ack/Nack from the SP router, a receiver should first verify if it has sent a PIM Switch Request or no. If not, it deletes the message. Otherwise, it handles the following tasks:

- If the message is a PIM Switch Ack, then the receiver has the permission to switch to the SPT. Therefore, it sends a PIM Join message toward the source.

[5]One could uses another policy to decide when to send the Switch Ack to the ISR. An alternative can be when the SP receives a Switch Accept from the majority of the oif list interfaces

- If the message is a PIM Switch Nack, the receiver can not switch to the SPT. It may try again after the expiration of [Switch-Request-Timer].

5.3 Illustration Example

We illustrate our mechanism using the example shown in Figure 3. The receiver R which detects that at least one of the parameter is violated (for example: the delay) sends a PIM Switch message towards the source S. The router Rt1 detects that it is the SP of the ISR. It then does not forward the switching request and it sends a PIM Switch Coordination message to interfaces 2 and 3 because they belongs to the *oif* list of the requested Group. We assume that receivers R_1 and R_2 decide to accept the switching coordination request. As consequence, they send a PIM Switch Accept message toward the SP. After receiving both of messages, the SP sends a PIM Switch Ack to the ISR. The ISR can then switch to the SPT by sending a PIM Join (S, G) toward the source.

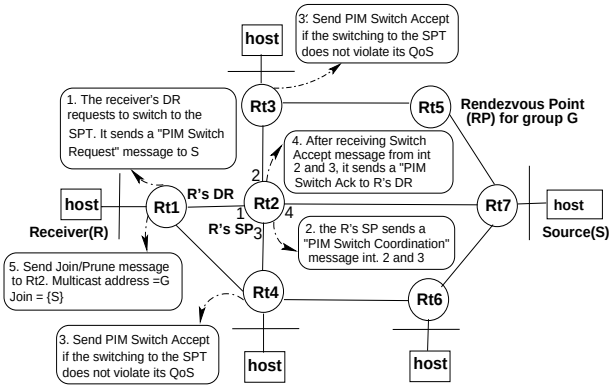

Figure 3: Example: Switching from shared tree (RPT) to shortest path tree (SPT). Actions are numbered in the order they occur

Now, assuming that receiver R_3 does not want to accept the switching coordination request, it then sends a PIM-SM Switch Refuse message toward the SP which in return sends a PIM-SM Switch Nack message toward the ISR without waiting for the response from the receiver R_4.

5.4 Eligibility tests

5.4.1 The case of delay constraint

We first consider the case when the QoS required is expressed in terms of an additive QoS parameter. Without loss of generality, we use the end-to-end delay as example, and impose an upper bound D_R, on the acceptable end-to-end delay perceived by a receiver router in a multicast group.

The receiver R maintains the required delay D_R and the current $D_R(RPT)$ of the multicast session received from the RPT. When detecting that the $D_R(RPT)$ remains superior to D_R for a configured period of time, it decides to switch to the SPT. It sends a switch request towards the source and it continues receiving data from the RPT while waiting for the response for its switch request. In addition to other information, this request should contain the values of D_R and $D_R(RPT)$. When receiving the switch request from the receiver R, the R's Switching Point (SP)[6] should decide if the switching to the SPT can improve the delay and if it degrades the QoS of other receivers belonging to the SP-based subtree.

Let us now detail how the the SP should process the switching request. The first thing that the SP should do is the estimation of the delay $D_R(SPT)$ denoting the delay if the data was received directly from the SPT. In fact, we have:

$$D_R(SPT) = D_R(RPT) - [D_{SP}(RPT) - D_{SP}(SPT)] \quad (9)$$

where $D_{SP}(RPT)$ and $D_{SP}(SPT)$ are the delay viewed by the SP when it receives data from the RPT and from the SPT, respectively. The only missing information to compute $D_R(SPT)$ is the value of $D_{SP}(SPT)$. We propose two manners to evaluate this delay:

- In the case when the routers between the SP and the source support a link-state unicast routing protocol such as OSPF, the SP can determine the delay from the source by cumulating the delays of the links in its path toward the source. In this case, we have:

$$D_{SP}(SPT) = \sum_{(i,j) \in P(SP,S)} d(i,j)$$

- In other cases (i.e., the unicast routing protocol does not provide this information), the SP can join directly the source for a short period of time and compute the delay and leave the source after this period[7].

After computing the value of $D_{SP}(SPT)$, the SP compares this value to the required value D_R. It considers that the SPT is more effective in term of delay than the RPT if:

$$D_R(SPT) \leq D_R. \quad (10)$$

The second eligibility tests that should be done by the SP is the assurance that the QoS requirements receivers in the SP-subtree are not violated due to the switching from the RPT to the SPT. In other words, it should verify that

[6] On the reception of a switch request, routers compare the outgoing interface to the source and the incoming interface from the RP. In the case when these interfaces are different, the PIM router is considered as the SP of that receiver.

[7] In addition of multicast packets sent to the RP encapsulated in PIM-SM Register Messages, the source unicast multicast packets to the SP during this period of time. A new PIM-SM message can be added to PIM-SM protocol in order to support this scenario.

the following equation holds for each concerned receiver X belonging to the SP-subtree:

$$D_X(SPT) \leq D_X + \varepsilon \quad (11)$$

where ε is an adjustment parameter estimated or configured by the PIM-SM router and D_X is the delay required by the receiver X. $D_X(SPT)$ is computed as follows:

$$D_X(SPT) = D_X(RPT) - [D_{SP}(RPT) - D_{SP}(SPT)] \quad (12)$$

When Eq. 10 holds and Eq. 11 holds for each concerned receiver, the SP assumes that the switching to the SPT is useful and it forwards the switching request to the source.

5.4.2 The case of rate constraint

When the receiver R detects that the equation $R_R(RPT) < R_R$ holds, it sends a switch request toward the source. In its request, the receiver specifies the value of the required rate (R_R) and the current received rate from the RPT ($R_R(RPT)$). When it receives the ISR's switching request, the R's SP compares $R_R(RPT)$ and $R_{SP}(RPT)$ denoting the rate received by the SP from the RPT. If the following equation holds:

$$R_R(RPT) < R_{SP}(RPT), \quad (13)$$

the bottleneck link belongs to the path $(SP,\ R)$ and not $(S,\ RP) \cup (RP,\ SP)$[8]. So, there is no need to switch from the RPT to the SPT since the rate will not be increased even if the rate $R_{SP}(SPT)$ is greater then $R_{SP}(RPT)$.

In the case when $R_R(RPT) = R_{SP}(RPT)$, we can conclude that the bottleneck link belongs to the path $(S,\ RP) \cup (RP,\ SP)$ and not $(SP,\ R)$. Then, the switching from the RPT to the SPT may increase the rate received by the receiver.

The missing information needed by the SP to take the switching decision is the value of $R_R(RPT)$ denoting the rate that can be received from the SPT and that of $R_{min}(SP, R)$ corresponding to the minimum available bandwidth in the path $(SP,\ R)$. We suppose that the SP gets these information by sending a request to PIM-routers belonging to the path toward the source (or by joining the source for a small period of time) and the receiver R (or via another manner, i.e., the use of a link-state unicast protocol), respectively.

After getting $R_{SP}(SPT)$ and $R_{min}(SP, R)$[9], the SP judges whether or not the switching to the SPT can improve the rate received by the receiver R. If:

$$R_{SP}(SPT) \geq R_R \text{ and } R_{min}(SP, R) \geq R_R \quad (14)$$

the receiver R will receive data from the SPT with the rate $R_R(SPT)$ given by:

$$R_R(SPT) = \min(R_{SP}(SPT),\ R_{min}(SP, R)) \geq R_R \quad (15)$$

The second task of the SP is to ensure that the QoS requirements of other switching concerned receivers in the SP-subtree are not violated when the receiver R switches from the RPT to the SPT. To do this, the SP can request the available minimum bandwidth $R_{min}(SP, R)$ in all paths of the SP-subtree toward the receivers and not only the path $(SP,\ R)$ and verify that:

$$R_{SP}(SPT) \geq \max_{R \in SP-subtree} R_R$$
and
$$R_{min}(SP, R) \geq \max_{R \in SP-subtree} R_R$$

for each receiver in the SP-subtree.

5.5 Implementation issues

The switching mechanism proposed in this paper can be easily integrated in PIM-SM protocol. Indeed, currently PIM-SM uses eight message type values among sixteen available values. We propose to add a new message type value (e.g., type number 9) to Switch messages. The sub-messages (Request, Coordination, Accept, Refuse, Ack, Nack) will be included in the switch message as indicated in Figure 4 which represents the Switch message format.

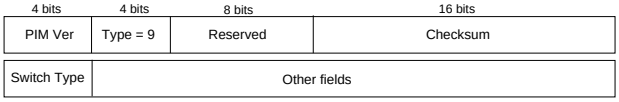

Figure 4: Switching message format

In a multicast delivery tree where there are on-tree routers that may implement different PIM-SM versions, i.e, some routers support the switching mechanism and others do not support it, as result we will the default behavior of PIM-SM. Indeed, for example when a receiver's DR does not respond to a PIM Switch Coordination message sent by the ISR's SP, the SP will behave as it receives a PIM Switch Accept. Then, it sends a PIM Switch Ack/Nack based only on the coordination between the downstream receivers that support our switching mechanism.

We choose the following values to the switch sub-types messages: Request message: type number 0, Coordination message: type number 1, Accept message: type number 2, Refuse message: type number 3, Ack message: type number 4, Nack message: type number 5.

[8]Considering that the rate is a concave parameter, the rate received by the ISR is the minimum of the available rates of links toward the RP or the Source. If this minimal belong to the path between the ISR and its SP, the rate received by the ISR from the RPT and the SPT is at least the same.

[9]The value of $R_{min}(SP, R)$ is superior to $R_R(RPT)$ because the bottleneck link belongs to the path $(S,\ RP) \cup (RP,\ SP)$ and not to $(SP,\ R)$.

Future extensions of our mechanism may add 11 new types given that we use 4 bits to encode the switch message type.

6 Performance Evaluation

In this section, we evaluate the performance of our switching mechanism using simulation.

6.1 Simulation Model

We use the Network Simulator [13] to evaluate the performance of our switching mechanism that we implement in NS simulator[10].

Considering that PIM-SM is an intra-domain routing protocol, we assume that a network topology belonging to an unique administrative domain. We generate 100 network topologies of 1000 nodes with different connectivities values using the Brite tool [8][11] which integrates the GT-ITM generator[12].

In our simulations, each group was assigned a single Rendez-vous Point (RP) which is randomly selected from a set of some centrally located nodes by a manner to approximately equilibrate the number of groups served by each RP.

The sources of each multicast group is assumed to generate traffic with a specific characteristics. We attribute to each multicast group a multicast application for which each receiver has specific QoS requirements in terms of delay, data rate, and loss probability. The values of these parameters are randomly generated for each receiver.

We distinguished several scenarios depending on the unicast routing protocol used: RIP, OSPF, or Euclidean-distance-based (EUC) and for different QoS requirement: delay, bandwidth, and delay and bandwidth.

6.2 Results and Observations

We performed simulations to compare the quality of the delivery tree built by PIM-SM when using our switching mechanism.

We first interest in the case when we do not use a coordination-based switching mechanism. That's mean that a receiver decides to switch to the SPT without collaborating with other receivers.

In a first experiment we evalute the fraction of unsatisfied receivers when a receiver belongs to the same delivery sub-tree decides to switch to the SPT. We conducted 500 simulation scenarios. In each scenario, we randomly choose the receiver initiating the switching.

In Figure 5, 6, 7 we plot this fraction when the QoS parameter is delay, bandwidth, and both delay and bandwidth, respectively. For each case, we consider the use of different unicast routing protocols: RIP, OSPF, and EUC.

As we can see from the plots, when the ISR decides to switch to the SPT, it affects the reception quality of other receivers. Indeed, for more than 80% of the conducted simulations, the fraction of unsatisfied receivers is more that 10% when we consider the bandaidth constraint with RIP protocol. We obtained similair results for other cases.

In Figure 8, we show the fraction of group members that are concerned by the switching request. As we can see, regardless the unicast routing protocol used, this fraction is variable and it reaches high values which are close to 100% for many simulation scenarios.

From the above results, we can conclude that the use of the PIM-recommended switching mechanism is not fair in the sense that when a receiver switches to the RPT, it affects the QoS of other receivers.

In a second experiment, we interest in the case when we use our switching mechanism to manage the switching from the SPT to the RPT. That is, the ISR's switching request will be accepted only when all the SP-subtree downstream receivers accept the switching. We varied the group size from 2 to 45 and for each case we conducted 100 simulation instances. For each multicast group, we use an uniform distribution to generate the location of the different receivers and sources in the network. We comput the average number of accepted switching requests. In Figure 9, we plot the fraction of accepted switching requests in function of group size for various unicast routing protocols and when the QoS requirement is the delay only, the bandwidth only, and the delay and the bandwidth.

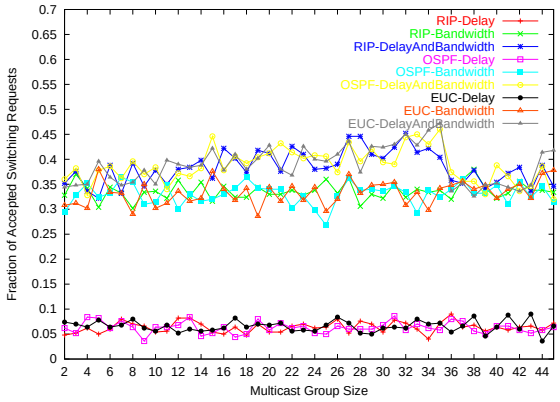

Figure 9: Variation of the fraction of accepted switching requests in function of the group size when using a coordination-based switching mechanism

[10]The tcl code of the different switching mechanisms is available at `http://www.inria.fr/rodeo/filali/pim-sm`.

[11]BRITE can be downloaded from `http://cs-www.bu.edu/brite/`.

[12]We have performed simulations with other generated topologies as well. The simulation results for these topologies are available at `http://www.inria.fr/planete/filali/pim-sm`.

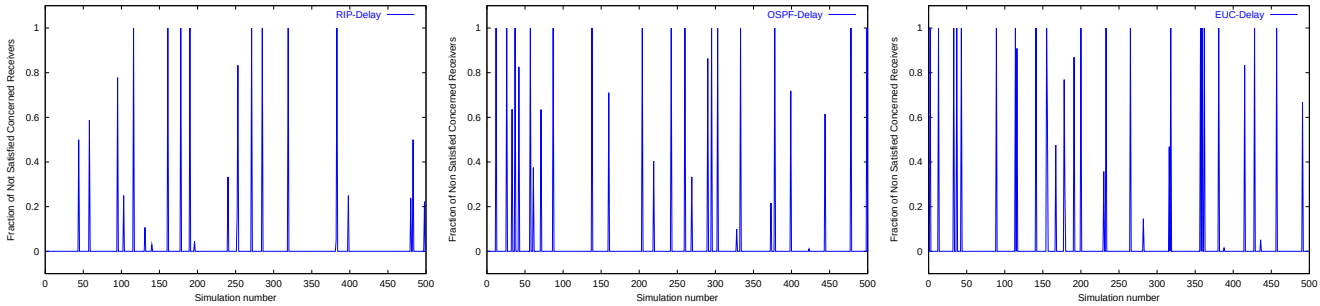

Figure 5: Fraction of unsatisfied receivers with delay constraint

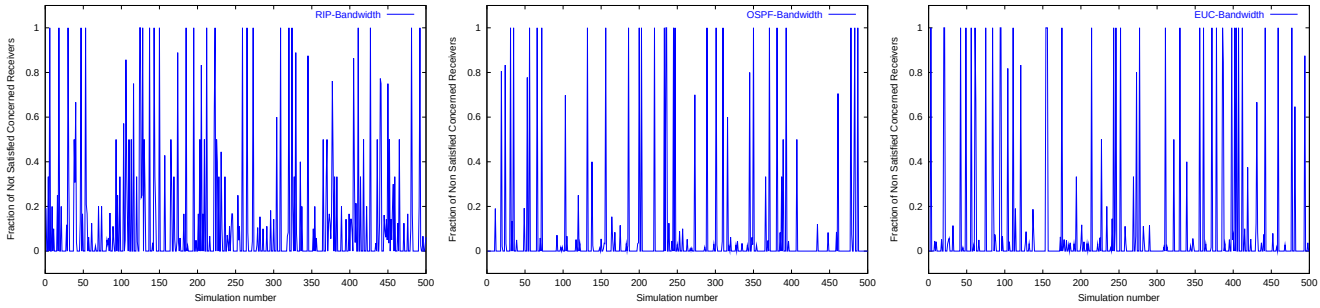

Figure 6: Fraction of unsatisfied receivers with bandwidth constraint

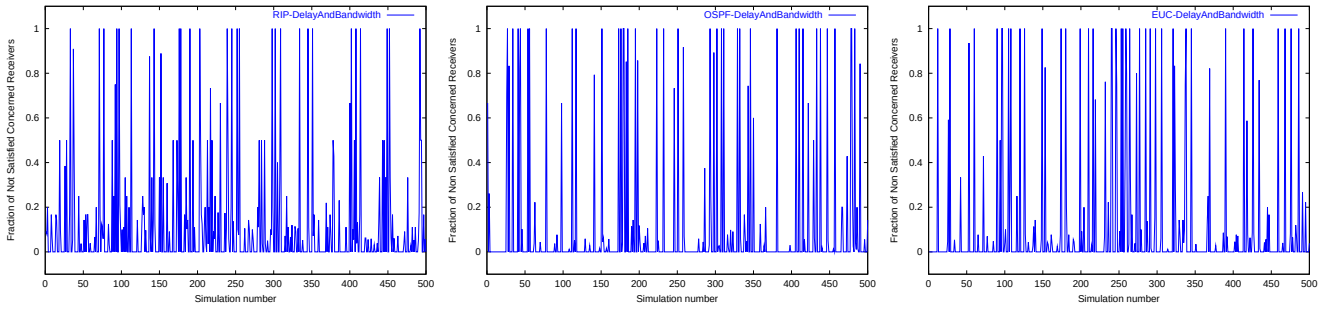

Figure 7: Fraction of unsatisfied receivers with delay and bandwidth constraints

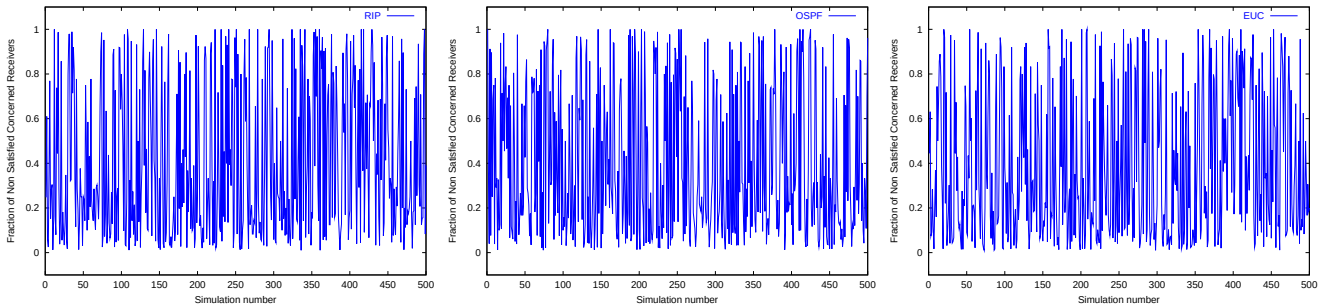

Figure 8: Fraction of the group members which are interested by the switching request

6.2.1 Switching Latency

We define the switching latency as the period of time between the time the receiver has sent the join request to the source and the reception of the join acknowledgment. The switching latency is always less than two times the timer [Switch-Coordination-Timer] plus the delay from the ISR to its SP.

We denote by T_l the latency of the switching receiver initiator. T_l can be computed as follows:

$$T_l = T_{ack/nack} - T_{request}$$

when the switching was initiated by a receiver.

The switching latency time depends on the complexity of the switching mechanism and the processing tasks that must be executed by each router before taking the switching decision.

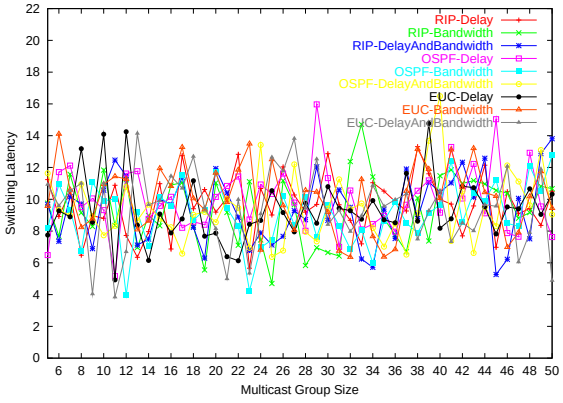

Figure 10: Variation of the switching latency

We plot in Figure 10, the variation of the switching latency in function of the group size. We can easily see that the switching latency varies between 4 ms and 16 ms. Regardless the unicast routing protocol used, curves have the same shape.

7 Conclusion and Open Issues

In this paper, we have addressed the problem of switching from the shared tree and the source-based tree in PIM-SM intra-domain multicast routing protocol. We demonstrated the necessity of a coordination-based switching mechanism to increase the effectiveness of PIM-SM. The switching can be initiated either a receiver's designated Router or the RP. We defined a new PIM-SM message (Switch Message) as well as sub-messages (Request, Coordination, Accept, Refuse, Ack, Nack) that can be easily included in the standard PIM-SM. Indeed, through the use of timers, the default behavior of our mechanism is that recommended by PIM-SM standard.

We used simulation to evaluate our proposed switching mechanism in the case of using RIP as well as OSPF as the underlying unicast routing protocol. The results show that our coordination-based switching mechanism improves the performance metrics of all receivers concerned by the switching decision. When combined with a QoS-based unicast routing protocol, our mechanism could perform much better then a coordination-less-based one as that described in PIM-SM standard [4].

Future work could evaluate other performance metrics of our switching mechanism such as the bandwidth overhead and the receivers satisfaction. Another area of future study is to investigate how our mechanism can extended to provide the switch back alternative to the RPT. We believe that the switch back can be integrated by the same way as our mechanism does.

References

[1] A. Ballardie, *Core Based Trees (CBT Version 2) Multicast Routing: Protocol Specification*, Internet Engineering Task Force, RFC 2189, September 1997.

[2] S. Deering and D. Cheriton, *PIM Architecture for wide-area multicast routing*, IEEE/ACM Transactions on Networking, pp. 153-162, April 1996.

[3] S. Deering, D. Estin, D. Farinacci, V. Jacobson, A. Helmy, D. Meyer, and L. Wei, *Protocol independent multicast version 2 dense mode specification*, IETF, draft-ietf-pim-v2-dm*.txt, June 1999.

[4] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, and L. Wei, *Protocol independent multicast sparse-mode (PIM-SM): Protocol specification*, IETF, RFC 2362, June 1998.

[5] F. Filali and W. Dabbous, *A PIM-SM Enhanced Switching Mechanim*, INRIA Research Report, April 2002.

[6] C. Hedrick, *Routing Information Protocol*, IETF, RFC 1058, June 1988.

[7] J. Holt and W. Peng, *Improving the PIM Routing Protocol with Adaptive Switching Mechanism between its Two Sparse Sub-Modes*, In the proceedings of the International Conference on Computer Communication and Networks, pp. 768-773, October 1998.

[8] A. Medina, A. Lakhina, I. Matta, and J. Byers, *BRITE: An Approach to Universal Topology Generation*, In Proceedings of the International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunications Systems - MASCOTS '01, Cincinnati, Ohio, August 2001.

[9] J. Moy, *OSPF Version 2*, IETF, RFC 2328, April 1998.

[10] M. Luby, V. Goyal, and S. Skaria, *Wave and Equation Based Rate Control: A massively scalable receiver driven congestion control protocol*, draft-ietf-rmt-bb-webrc-00.txt, October 2001.

[11] J. Widmer and M. Handley, *TCP-Friendly Multicast Congestion Control (TFMCC): Protocol Specification*, draft-ietf-rmt-bb-tfmcc-00.txt, November 2001.

[12] L.Wei and D˙Estrin, *Multicast Routing in Dense and sparse Modes: Simultion Study of Tradeoffs and Dynamics*, In the proceedings of the 4^{th} International Conference on Computer Communications and Networks (ICCCN), September 1995.

[13] Network Simulator `http://mesh.cs.berkley.edu/ns/`