A survey of multicast protocols for Internet*

J. Ignacio Alvarez-Hamelin† and Pierre Fraigniaud‡

Soumission AlgoTel 2000
Correspondant : I. Alvarez-Hamelin

Abstract

The aim of this paper is to survey the main multicast protocols that were proposed for Internet. We classify these protocols according to several parameters, including the way group-members are connected, the way the members are aggregated to the current group topology, the scale of the protocol (inter or extra domain), the possibility to take into account specific QoS requirements, etc. The purpose of this paper is not to be exhaustive, but rather to describe the main characteristics of the current multicast protocols.

1 Introduction

As opposed to unicast protocols connecting pairs of users, multicast protocols [4, 9] support group communications, e.g., white board, video-on-demand, and video-conferences, in which more than two users, possibly hundreds of users, interact. In most of the existing protocols, the structure of the multicast group is a tree connecting all the members, and messages exchanged among the group members are routed along the links of that tree.

Several methods have been proposed for constructing multicast trees offering a good cost vs. performance ratio. The main difficulty comes from the dynamic of the group in the sense that members can join and leave at their convenience. On one side, many algorithms have been derived to solve the *Steiner* tree problem. This problem is NP-complete but efficient $O(1)$-approximation algorithms have been derived. The main problem of this approach is to be centralized and off-line. One the other side, on-line decentralized protocols have been derived. The main problem of the on-line approach is there exists worse case sequences of join-leave that make the resulting tree far from the optimal [7]. However, the decentralized nature of Internet forces the use of on-line protocols.

One can find dozen of (decentralized) multicast protocols in the literature. The aim of this paper is to answer why so many protocols were introduced. For that purpose, we give a brief description of some of the most famous multicast protocols, and, throughout this description, we show the main evolution of the protocols from the early 90's to nowadays. The results of this study are summarized in Table 1.

In our description of the protocols, the following parameters are taken into account. (These are the parameters appearing as column titles in Table 1.)

Type of connection. Roughly speaking, multicast protocols are either based on a single tree shared by all the members, or based on several trees, e.g., one for each source of the group. That is, we make a distinction between *shared* tree and *source-based* tree.

Aggregation rule. The aggregation is said *greedy* if a joining node connects to the closest node already in the group. The aggregation is said *RSP* (Reverse Shortest Path) if a joining node is connected to the group by a shortest path to the source(s).

*Supported by the program A.R.P. of the CNRS, and by the French RNRT program "ROM".

†Laboratoire de Recherche en Informatique, Université Paris-Sud, 91405 Orsay cedex, France. `ihameli@lri.fr`. Also: Facultad de Ingeniería, Universidad de Buenos Aires, Paseo Colón 850, (1063) Buenos Aires, Argentina. Additional support from the *Alpha* program.

‡Laboratoire de Recherche en Informatique, Université Paris-Sud, 91405 Orsay cedex, France. `pierre@lri.fr`. Additional support from the CNRS.

	Connection	Aggregation	QoS	Construction	Updates	Loop	Domain
DVMRP	Source-based	RSP	No	Broadcast/Pruning	Periodic	-	Intra
MOSPF	Source-based	RSP	No	OSPF + Group	Unicast-driven	-	Intra
CBT	Shared	RSP	No	Unicast	No	-	Both
PIM-SM	Shrd & S-B	RSP	No	Unicast/Pruning	Shrd $\rightarrow$ S-B	-	Both
YAM	Shared	Greedy	Yes	Multiple paths	No	-	Both
QoSMIC	Shrd & S-B	Greedy	Yes	Multiple paths	Shrd $\rightarrow$ S-B	No	Both
MIP	Shrd & S-B	SP & Grdy	No	Breadth-first search	Unicast-driven	No	Both
BGMP	Shared	RSP	No	BGP	Unicast-driven	-	Inter
SM	Shared	RSP	No	Unicast	Unicast-driven	No	Both

Table 1: Properties of multicast protocols

Quality of service. Some protocols aim to optimize parameters such as delay, bandwidth, etc. Although one cannot formally talk about QoS parameters, some protocols are more concerned by certain aspects of QoS than others.

Construction. There are many techniques for constructing the structure connecting the group members. Many of them are based on the underlying unicast protocol (e.g., OSPF, BGP). Some others use BFS techniques. Others use a pruned spanning tree of the whole network. Others explore multiple paths and keep the "best" path.

Updates. There are mostly two reasons for updating a tree. One is that the several members joining and leaving the group modify the structure of that group significantly. The other is that the network conditions (traffic, connection, etc.) may have changed. Some multicast protocols refresh the tree(s), either periodically or based on specific events (often generated by the underlying unicast protocol).

Loop. Some protocols are proved loop-free. For the others, either there exist situations for which loops can occur, or such situations does not exist but this is not formally proved.

Domain. A multicast protocol can be implemented at two scales: inter or extra domain. Inter domain, a protocol can make use of broadcasting, whereas this is not allowed at the extra-domain scale.

2 Survey of multicast protocols

This section gives a brief description of some of the multicast protocols previously proposed in the literature. We have listed these protocols in an order that somewhat reflects their date of birth.

2.1 Distance Vector Multicast Routing Protocol (DVMRP)

DVMRP [14] is among the first protocols for multicast in Internet. It does not assume specific unicast protocol. Its name comes from the use of the *Distance Vector* distributed algorithm to construct the trees. Every source constructs its own tree using a *broadcast and prune* technique as follows.

Intra-domain, the source s broadcasts a message M. Upon reception through interface i of a message M originated at s, every router r does the following. (The interface i is called the *upstream* interface, whereas the other interfaces are called *downstream* interfaces.) Node r either forwards M through all its not yet pruned downstream interfaces, or discards M and sends a prune message back through the upstream interface i if all its downstream interfaces are pruned. An interface connected to a LAN containing no member of the group is pruned. When a new member joins the group, it removes the prune filters along its path to the source. Periodically, the prune filters are removed, so that the multicast protocol can takes into account possible changes of the unicast topology.

Inter-domain, DVMRP uses *tunnels* between domains. The protocol then works upon these tunnels (to avoid broadcast over the whole Internet). A tunnel between two domains is set by the administrators of the two domains. The global topology of the tunnels result from negotiations between the administrators of the domains.

2.2 Multicast Open Shortest Path First (MOSPF)

MOSPF [10] is an intra-domain multicast protocol that constructs one tree per source in every group. It is an extension of OSPF (*link-state* routing), that is every node must be aware of the whole network topology and group memberships. MOSPF aims to construct the tree of source s by merging the shortest paths from s to the members. For that purpose, every node joining the group broadcasts its request to the whole domain. Every router then updates its multicast tables by application of an algorithm common to all nodes (every router knows the entire topology, the several sources of the group, the addresses of the members, etc.). Tie-breakers have been defined to ensure that, when several equal-cost paths exist, all routers agree on which single path to use.

2.3 Core-Based Tree (CBT)

CBT [1] was among the first proposed protocols using a single tree shared by all the members of the group. This tree is rooted at a specific node, called the *core* of the group. Any member aiming to broadcast a data to the group, sends this data to the core which initiates the broadcast along the edges of the tree. Although the length of the routes from the source to the destinations can be twice as long as the routes in a shortest paths tree rooted at the source, the gain in the size of the routing tables balances this drawback. Actually, most of the proposed multicast protocols which came after CBT make use of a shared tree, mainly because managing a single tree is much easier that controlling many trees.

CBT yields many theoretical and practical problems. E.g., what is the optimal choice for the core? At which frequency should we recompute the position of the core to take into account the possible modification of the group? What is the tradeoff between the number of cores, and the efficiency of the multicast? etc. Some of these problems are addressed in [6].

The aggregation of the nodes follows the RSP technique as follows. Every node x of a LAN aiming to join some group G informs the gateway of the LAN. This router will then send a request to the first node along the path toward the core of G, and will add an entry for x in its multicast table. Every router along the path from x to the core propagates in turn the join request of x, and adds an entry for x in its multicast table. The construction of the tree is therefore simply based on the underlying unicast protocol which is in charge of finding the shortest paths from the members to the core. The relative simplicity of CBT allows it to perform intra- and inter-domain as well.

2.4 Protocol Independent Multicast - Sparse Mode (PIM-SM)

PIM-SM [3] was defined almost as the same time as CBT. It is based on roughly the same technique (the core is called *rendez-vous point*). The main difference between PIM and CBT is the ability of PIM-SM to move from a shared-tree topology to a source-based trees topology if some sources deliver high rate traffic. This change is actually driven by the gateways of the group members. If a gateway x detects that some source s generates high traffic, x connects to s in the way it connected to the rendez-vous point. The gateway x also sends a prune message to the first node along the path toward the rendez-vous point to inform that router that it must not forward messages from s to x on the shared-tree anymore. This pruning is propagated upward the shared tree in roughly the same way as it is done in DVMRP.

Although PIM-SM was designed for both inter- and intra-domain, there are many discussions about the scalability of this protocol because of the relative difficulty of managing many trees for every group.

2.5 Yet Another Multicast protocol (YAM)

YAM [2] is based on a shared tree topology. The aggregation process is greedy as opposed to the RSP technique used in PIM or CBT. More precisely, when a new member x joins the group, it is connected to the tree by a shortest path from x to the tree. The node of the tree that is closest from x is found by a local search performed at incremental distances, or by a broadcast limited to nodes at a given distance from x.

YAM actually aims to support QoS in the following sense. YAM looks for multiple alternative paths for connecting a new member to the current tree. The chosen path connecting a new member to the tree is then not necessarily the shortest, but can also be selected according to some QoS parameters, e.g., link capacity, delay, reliability or cost, if routers support resource reservation protocols (e.g., RSVP [13]

For inter-domain multicast, YAM assumes that every domain has a specific node called *egress*. A new member x contacts the egress node of its domain. The egress then broadcasts on the *well-known multicast address* of YAM to look for the current tree. (A well-known address is specified in advance, and is specific of the protocol, here YAM.) Again, the best egress-to-egress path is selected among the several choices according to some QoS parameters.

2.6 Quality of Service sensitive Multicast Internet protoCol (QoSMIC)

As PIM-SM, QoSMIC [5] is able to support simultaneously a shared tree, and several source-based trees for sources generating high traffic. As YAM, QoSMIC is aimed to support QoS. The aggregation technique of QoSMIC is roughly the same as the aggregation technique of YAM. However, QoSMIC also offers the possibility for a new member x to contact a *manager* in charge of returning to x the nodes of the current tree(s) to which x can connect. Node x then contacts these nodes to get information of the QoS along the path connecting x to them, in order to select the best path. This ability allows QoSMIC to perform in intra- and inter-domain as well.

2.7 Multicast Internet Protocol (MIP)

MIP [11] was designed aiming to overcome some limitations of both CBT (e.g., shared tree) and PIM-SM (e.g., control messages). As a consequence, MIP allows the use of a shared tree with few source-based trees for sources generating high traffic. The dynamic evolution of the tree(s) is driven by the evolution of the network. MIP offers two modes for the construction of the tree(s): *sender-initiated* and *receiver-initiated*. The former is on shortest path (not on RSP), whereas the latter is greedy.

The sender-initiated protocol can be initiated by a source that wants to make use of a personal tree to broadcast its messages to the group. As opposed to PIM-SM, it is not the initiative of a member to create a new tree. Roughly speaking, the sender-initiated protocol constructs a breadth-first search tree (shortest path tree) spanning the group members. Actually, MIP allows a certain tolerance between the shortest path distance in the tree(s) and the shortest path distance in the network in order to reduce the frequency of modification of the tree(s).

The sender-initiated protocol makes a significant difference between MIP and the previously listed protocols, which all construct the tree(s) in a member-driven manner.

2.8 Border Gateway Multicast Protocol (BGMP)

BGMP [8] is an *inter*-domain multicast protocol. It is in charge of constructing an inter-domain shared-tree based on BGP protocol for inter-domain unicast (*Border Gateway Protocol*). The multicast tree is built using the unicast BGP information, and hence the aggregation process is RSP.

The goal of BGMP not only to extend BGP, but also to assign multicast addresses. There were other solutions for assigning multicast addresses, but they were not scalable. BGMP assigns the addresses in a hierarchical manner through the use of the MASC (*Multicast Address-Set Claim*) protocol which is in charge of maintaining dynamically a set of multicast address between all the domains, and the protocol *listen and claim with detection approach* which controls requests for addresses using a logical tree connecting the autonomous systems.

2.9 Simple Multicast (SM)

SM [12] uses a shared-tree, in a similar way as CBT does. The goals of SM is mainly to provide the ability of using a unique protocol for both inter- and intra-domain. Indeed, as we already said, there are some scalability problems in the previously listed inter-domain protocols (apart BGMP), mainly due to the allocation of the multicast addresses. SM proposes the use of the pair (address of the core, address of the multicast group) as unique identification of the multicast group. In this way, there is no need to ask to the whole Internet whether there is a core handling a multicast group with the same address.

It is worth to note that SM has also an explicit loop detection.

3 Conclusion

In this paper, we have surveyed some of the most famous protocols for multicast in the Internet. We have focussed our attention on the global structure of these protocols, i.e., tree topology, aggregation process, etc, although these protocols may also differ by other important parameters. Indeed, the purpose of the authors (and they are currently working on it) is to propose protocols that should allow to speed-up the communications of existing multicast protocols via an appropriate scheduling of the point-to-point communications, and/or via the construction of trees using the links with minimum congestion.

References

[1] T. Ballardie, P. Francis, and J. Crowcroft. Core based tree (CBT): an architecture for scalable inter-domain multicast routing. In *proceedings of ACM SIGCOM*, pages 85–95. ACM press, 1993.

[2] K. Carlberg and J. Crowcroft. Building shared trees using a one-to-many joining mechanism. In *ACM SIGCOMM Computer Communication Review*, pages 5–11, January 1997.

[3] S. Deering, D. Estrin, D. Farinacci, V. Jacobson, C-G. Liu, and L. Wei. An architecture for wide-area multicast routing. In *SIGCOM*, pages 126–135. ACM press, 1994.

[4] C. Diot, W. Dabbous, and J. Crowcroft. Multipoint communication: a survey of protocols, functions, and mechanisms. *IEEE journal on selected areas in communications*, 15(3):277–290, 1997.

[5] M. Faloutsos, A. Banerjea, and R. Pankaj. Qosmic: Quality of service sensitive multicast internet protocol. In *SIGCOMM*, September 1998. Vancouver BC.

[6] E. Fleury, Y. Huang, and P. K. McKinley. On the performance and feasibility of multicast core selection heuristics. In *Seventh International Conference on Computer Communications and Networks*, October 1998. Lafayette, Louisiana.

[7] P. Fraigniaud and S. Vial. Heuristic Algorithms for Personalized Communication Problems. In *SIROCCO '97*, pages 240–252, 1997.

[8] Satish Kumar, David Thaler, Cengiz Alaettinoglu, Deborah Estrin, and Mark Handley. The masc/bgmp architecture for inter-domain multicast routing. In *ACM SIGCOMM '98*, September 1998.

[9] W. Mostafa and M. Singhal. A taxonomy of multicast protocols for internet applications. *Computer communications*, 20:1448–1457, 1998.

[10] J. Moy. Mospf: Analysis and eperience RFC 1595. Technical report, IETF, 1994.

[11] M. Parsa and J.J Garcia-Luna-Aceves. A protocol for scalable loop-free multicast routing. *IEEE Journal on Selected Areas in Communications*, 15(3):316–331, April 1997.

[12] R. Perlman, A. Ballardie, J. Crowcroft, Z. Wang, T. Maufer, C. Diot, J. Thoo, and M. Green. Simple multicast: A design for simple, low-overhead multicast Internet Draft draft-perlman-simple-multicast-03.txt. Technical report, IETF, October 1999.

[13] M. Pullen, R. Malghan, L. Lavu, G. Duan, J. Ma, and H. Nah. A simulation model for ip multicast with rsvp RFC 2490. Technical report, IETF, Jaunary 1999.

[14] T. Pusateri. Distance vector multicast routing protocol Internet Draft draft-ietf-idmr-dvmrpv309.txt. Technical report, IETF, September 1999.