Scalable IP multicast for many very small groups with many senders and its application to mobility

Ljubica Blazević and Jean-Yves Le Boudec

Institute for computer Communications and Applications (ICA)
Swiss Federal Institute of Technology, Lausanne
email: {Ljubica.Blazevic, Leboudec}@epfl.ch

Abstract

We consider the problem of multicast routing in a large single domain network with a very large number of multicast groups with small number of receivers. Such a case occurs, for example, when multicast addresses are statically allocated to mobile terminals, as a mechanism to manage Internet host mobility [7]. For such networks, existing dense or sparse mode multicast routing algorithms do not scale well with the number of multicast groups.

We propose an alternative solution called Distributed Core Multicast (DCM) that is based on an extension of the centre-based tree approach. We also describe how our approach can be used to support mobile terminals.

Introduction

We present a solution for providing low overhead delivery of multicast data in a large single domain network for a very large number of small groups. Such a case occurs when the number of multicast groups is very large (for example, greater than a million), the number of receivers per multicast group is very small (for example, less than five) and each host is a potential sender to a multicast group. We propose to apply this solution to support mobility in the Internet where a multicast address is statically assigned to a mobile host.

MSP-IP (Mobility support using Multicasting in IP)[7] proposes a generic architecture to support host mobility in the Internet by using multicasting as the mechanism for routing packets to the mobile hosts. Every mobile host is statically assigned and addressed by a multicast address. A multicast router in a mobile host's current cell is responsible for joining the multicast distribution tree on behalf of a mobile host. This multicast router typically coexists with the base station in a cell. A base station that anticipates the arrival of a mobile host initiates a mobile host's group membership registration. Thus, a multicast group assigned to a mobile host has a few recipients. At the same time, a mobile host receives data only from a base station in its current cell. Hence, we have a form of unicast end-to-end communication which uses multicast routing. See [7] for a detailed description of the implications of using multicast addresses to support mobile hosts.

The benefits of using multicast addresses to support mobile IP are twofold:

- A fixed multicast address is assigned to a mobile host. This simplifies the task of the correspondent host and eliminates the need of explicit address translation (as in other proposals: IETF Mobile IP [8], SONY Scheme [9], IPv6 Mobility Proposal [3]).

- In the proposals mentioned above, a response to a handoff of the mobile host happens only after the home agent (correspondent host in IPv6) becomes aware of the host's new location. In contrast, when a multicast address is assigned to the mobile host, base stations in neighbouring cells could have already joined the multicast group assigned to a mobile host through advance registration. This minimises packet losses and latency when a mobile host changes its location.

We propose an extension to an existing multicast routing protocol which aims to scale better for the design objectives mentioned above (many large groups with very few senders). Recent sparse multicast routing protocol efforts, such as the protocol independent multicast (PIM-SM) [5] and the core-based trees (CBT) [2], build a single delivery tree per multicast group which is shared by all group's senders. This tree is rooted at the single centre router. In CBT, the centre is called the "core". In PIM, a group-shared tree is rooted at a rendezvous point (RP).

Centre-based routing protocols have potential shortcomings:

- Traffic for the multicast group is concentrated on the links along the shared tree and particularly near the core router.

- Finding an optimal centre for a group is a NP-complete problem and requires the knowledge of the whole topology [12]. Current approaches typically use either administrative selection of centres or some simple heuristics [10]. Data distribution through a single core router could cause non optimal distribution of traffic in the case of a bad positioning of the core (or the RP) router with respect to senders and receivers. This problem is known as a triangular routing problem.

We propose an alternative solution, called Distributed Core Multicast (DCM), for the efficient and scalable delivery of multicast data under the assumptions that are satisfied when multicast is used to support mobile IP (large number of multicast groups, a few receivers per group and a potentially large number of senders to a multicast group). We consider a network model that consists of several areas connected via the backbone area (see Figure 1). The objectives we want to achieve are: (1): avoid state information in backbone routers, (2): avoid triangular routing across expensive backbone links and (3) scale well with the number of multicast groups. Our solution is based on an extension of the centre-based tree approach.

The following is a short description of our proposal. We introduce an architecture based on several core routers per multicast group, called Distributed Core Routers (DCRs). The DCRs in each area are located at the edge of the backbone. The DCRs act as backbone access points for the data sent by senders inside their area to receivers outside this area. A DCR also forwards the multicast data received from the backbone to receivers in the area it belongs to. When a host wants to join the multicast group m, it sends a *join* message. This *join* message is propagated hop-by-hop to the DCR inside its area that serves the multicast address. Conversely, when a sender has data to send to the multicast group, it will send the data encapsulated to the DCR assigned to the multicast address.

The Membership Distribution Protocol (MDP) runs between the DCRs serving the same range of multicast addresses. It is fully distributed. MDP enables the DCRs to learn about other DCRs that have group members.

Distribution of data uses a special mechanism between the DCRs in the backbone area, and the trees rooted at the DCRs towards members of the group in the other areas. We propose a special mechanism for data distribution between the DCRs that does not assume that non-DCR backbone routers perform multicast routing. We propose an initial solution to this mechanism that can be applied today. The final solution is the object of ongoing work.

With the introduction of the DCRs close to any sender and receivers, converging traffic is not sent to a single centre router in the network. Data sent from a sender to a group within the same area is not forwarded to the backbone. Our approach alleviates triangular routing problem common to all centre-based trees. The DCM approach is implemented by using Network Simulator (NS) tool [1]. We have examined the properties of the DCR approach in a large single domain network. However, the DCM approach is not constrained to one domain network. Our future work would be to examine interoperability of DCM with other inter-domain routing protocols.

This paper is organised as follows. In the next section, we give a detailed description of our approach for scalable delivery of multicast data. Then, we evaluate the applicability of our approach to support IP host mobility. Finally, we give directions for future work and conclude the paper.

Description of the DCM approach

In this section, we describes the various elements of the approach. Those are: addressing issues, how members join the multicast group, how a sender sends to a multicast group, how membership information is distributed between DCRs and lastly, how multicast data is distributed between DCRs.

In order to describe the DCR approach, we use the network model that is presented on Figure 1.

Addressing Issues

In each area there are several routers that are configured to act as candidate DCRs. The identities of candidate DCRs are known to all routers within an area by means of a intra-area bootstrap protocol [4]. This is similar to PIM-SM with the difference that the bootstrap protocol is constrained within an area. This entails periodic distribution of a set of reachable candidate DCRs to all routers within an area. Routers use a common hash function to map any multicast address to one router from the set of candidate DCRs.

For a particular group M, we use the hash function to determine a DCR that serves M. The used hash function is $h(r(M), DCR_i)$. This function is similar to that used in Cache Array Routing Protocol (CARP)[11]. In our approach, function $r(M)$ takes as input the multicast group and gives as output the range of the multicast group, while DCR_i is the DCR address. The target DCR_i is then chosen as the one with the highest value of $h(r(M), DCR_j))$ among all j's from set $\{1, .., N\}$ where N is the number of candidate DCRs in an area.

One example of the function that gives the range of the multicast address M:

$$r(M) = M\&S \text{ ,where S is a bit mask.}$$

Each candidate DCR is aware of all ranges of multicast addresses for which it is elected to be a DCR. There is a function $m(r(M))$ that maps the range of the multicast address M to another control multicast address. A DCR joins a control multicast address that corresponds to a range of multicast addresses that it serves. This multicast address is used by the DCRs in different areas that serve the same range of multicast addresses to exchange control messages. All routers in all non-backbone areas should apply the same functions $h(..), r(..)$ and $m(..)$.

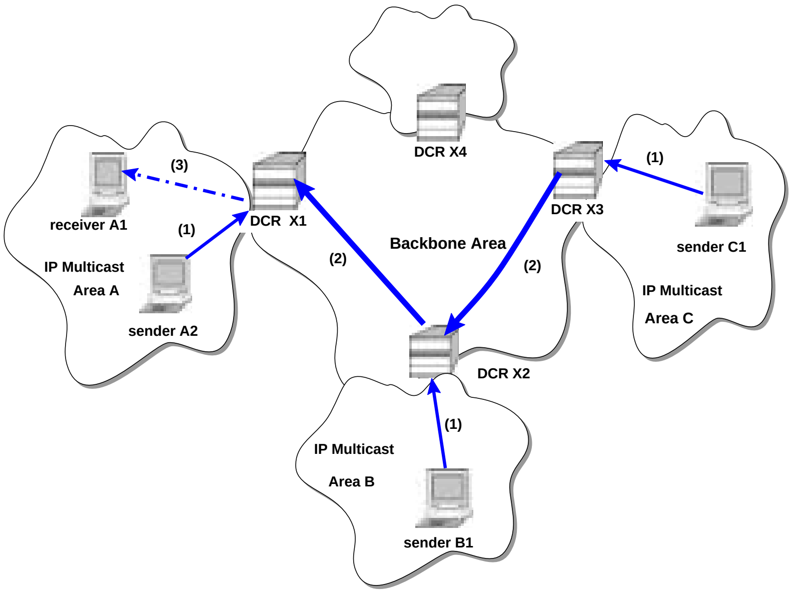

Figure 1: Model of a large single domain network and an overview of data distribution with the DCM approach. We show one multicast group m and DCRs X1, X2, X3 and X4 that serve a range to which m belongs. Step (1): Senders A2, B1 and C1 send data to the corresponding DCRs inside their areas. Step (2): DCRs distribute the multicast data across the backbone area to DCR X1 that needs it. Step (3): A local DCR sends data to the local receivers in its area.

How members join the multicast group

When a host wants to join a multicast group M, it issues a join via the IGMP. A router on its LAN known as designated router (DR) receives the IGMP join message and determines the DCR inside its area that serves that multicast address. Now, the process of the establishing of the group shared tree is like in PIM-SM [5]. The DR sends a *join* message towards the determined DCR that serves the multicast group. Sending a *join* message forces any off-tree routers on the path to the DCR to forward a *join* message and join the tree. Each router on the way to the DCR keeps a forwarding state for a multicast group M. When a *join* message reaches the DCR, this DCR now becomes labelled with the multicast address M. In this way, the delivery subtree for receivers of the multicast group M in an area is established. The subtree is maintained by periodically refreshing the state information for the multicast group M in the routers (like in PIM-SM, this is done by periodically sending *join* messages).

Like in PIM-SM, a DR that no longer has receivers for the multicast group sends a prune message towards the nearest DCR to disconnect from the shared distribution tree.

How senders send to a multicast group

The sending host originates native multicast data for the multicast group *M* that is received by the designated router (DR) on its LAN. The DR determines the DCR within its area that serves the multicast address. The DR encapsulates the multicast data packet (IP-in-IP) and sends it with the destination address the same as the address of the determined DCR. This DCR, referred to here as a source DCR, receives the encapsulated multicast data. This is similar to PIM-SM where the DR sends encapsulated multicast data to the RP corresponding to the multicast group. When the source DCR receives multicast data from a sender within its area, it distributes the data to members of the multicast group M inside its area and to other DCRs that have receivers for M in their corresponding areas. The source DCR delivers multicast data to local members in its area by sending the data along the preestablished subtree for *M*. In order to send multicast data to receivers in other areas, a source DCR needs to know the list of the DCRs in other areas that are labelled with the multicast address (those are the DCRs that have local receivers in their areas). The next subsection presents how the source DCR learns about the list of labelled routers.

How membership information is distributed between DCRs

The Membership Distribution Protocol (MDP) is used by the DCRs to exchange control information. All DCRs that serve the same range of multicast addresses within their corresponding areas are members of a MDP control multicast group. An MDP control multicast address is used for sending MDP control messages. Maintaining of the multicast tree for the MDP control multicast group is done by means of some existing multicast routing protocol (e.g CBT). A DCR joins as many MDP control multicast groups as there are ranges of multicast addresses that it serves.

The DCRs that serve the same range of multicast addresses in various areas are aware of each other. This is done by every DCR router periodically sending keepalive messages to the corresponding MDP control multicast address.

For the purpose of the distribution of the multicast data between the DCRs (as described below) each DCR router sends to the MDP control multicast address information about the unicast distance from itself to other DCRs that it learns to serve the same range of multicast addresses. This information comes from existing unicast routing tables.

In addition, a DCR, that is labelled with a multicast group *M*, informs all other DCRs that are responsible for the same multicast group that it has receivers for the multicast group M. The goal is that each DCR keeps a record of every other DCR which has at least one member for a multicast address from the range that the DCR serves. Since hosts can dynamically join and leave a multicast group, a DCR router needs to inform periodically all other DCR routers if it has receivers listening to the multicast group *M*. As soon as a DCR stops being labelled with some multicast address, it should inform all other DCRs.

In our approach, DCRs join all MDP control multicast addresses for all ranges that they can potentially serve. It is possible to reduce the number of MDP control multicast addresses that a DCR subscribes to by using some heuristics (this is the object of ongoing work).

Also, the approach presented here for MDP uses MDP control multicast addresses and flooding inside the groups defined by those addresses. An alternative approach would be to use MDP servers. This would be more complex, but also more scalable. This approach is not studied in detail in this paper.

How multicast data is distributed between DCRs

The multicast data for a group *M* should be distributed from a source DCR to all DCRs that are labelled with *M*. Since we assume that the number of receivers per multicast group is not large, there are only a few labelled routers per multicast group. Our goal is to perform multicast data distribution in the backbone in such a way that backbone routers keep minimal state information while at the same time backbone bandwidth is used efficiently. We propose a solution to perform data distribution between the DCRs that can be applied today.

Point-to-Point Tunnels In order to distribute multicast data from the source DCR to a list of labelled DCRs for the multicast group, an overlay solution can be applied by means of automatic tunnels.

This solution consists in that the source DCR calculates the tunnel tree that spans itself and labelled DCRs for some multicast address. The problem of spanning a subset of nodes in the graph with the minimal cost is known as the Steiner tree problem. We propose that the source DCR applies some heuristics.

Point-to-point tunnels are used to carry multicast data between the DCRs. Inter-DCR distribution information is put by the source DCR as an explicit distribution list in the end-to-end option field of the packet header. Under the assumption that the number of labelled DCRs for a multicast group is small, the number of labelled DCRs for that group should also be small. Thus, an explicit distribution list is not expected to be long.

When a DCR router receives a packet from another DCR, it reads from the distribution list whether it should make copies of multicast data and identities of the DCR routers where it should send encapsulated multicast data. Labelled DCRs deliver data to local receivers in the corresponding area.

Using point-to-point tunnels is a simple solution. The drawback is that backbone bandwidth is not used optimally because of packet duplications.

How to apply the DCM approach to support host mobility

In this section we present how the DCM approach can be used as a mechanism for routing packets to the mobile hosts.

We start this section with a short description of certain existing proposals for providing mobility in the Internet and then illustrate how the DCM approach can support mobility.

Overview of proposals for providing mobility in the Internet

In the IETF Mobile IP proposal [8] each host has a permanent home IP address that does not change regardless of the mobile host's current location. When the mobile host visits a foreign network, it is associated with a *care-of address* that is IP address related with the mobile host current position in the Internet. When a host moves to visited network it registers its new location with its home agent. The home agent is a machine that acts as a proxy on behalf of the mobile host when it is absent. When some stationary host sends packets for the mobile host it addresses them to the mobile host's home address. When packets arrive on the mobile host's home network, the home agent intercepts them and sends by encapsulation packets towards the mobile host's current location. With this approach all datagrams addressed to a mobile host are always routed via its home agent. This causes so-called triangle routing problem. Delivering packets in the opposite direction, from the mobile host to the stationary host, is straightforward. The mobile host sends directly to the stationary host, but the source address field in the IP packet is set to the mobile host's home address.

In the IPv6 mobility proposal [3] when a handoff is performed, the mobile host is responsible for informing its correspondent hosts about its new location. This is done by sending binding updates. Sending to a mobile host is done via a home agent in case where a correspondent host does not have binding for a mobile host.

The Columbia approach [6] was designed to support *intra*campus mobility. Each mobile host always retains one IP home address, regardless of where it is on the network. There is a number of dedicated Mobile Support Stations (MSSs) that are used to assure the mobile host's reachability. Each mobile host is always reachable via one of the MSSs. When a mobile host changes its location it has to register with a new MSS. A MSS is thus aware of all registered mobile hosts in its wireless cell. A source that wants to send a packet to a mobile host sends it to the MSS that is closest to the source host. This MSS is responsible for learning about the MSS that is closest to the mobile host and to deliver the packet. A special protocol is used to exchange information among MSSs.

MSM-IP (Mobility support using Multicasting in IP) [7] proposes a generic architecture to support host mobility in the Internet by using multicasting as a mechanism to route packets to the mobile hosts. Hence, there is a need for a new efficient multicast routing protocol to support host mobility.

The DCM application to host mobility

The DCM is therefore designed as a multicast routing protocol to support host mobility where each mobile host is statically assigned one class-D multicast address.

The routing of packets destined to the mobile host is done as it is described in the section that presented the DCM approach. For the mobile host's assigned multicast address, within each area, there exists a DCR that serves that multicast address. Those DCRs are responsible for forwarding data to a mobile host. As was said before, the DCRs run MDP control protocol and are members of a MDP control multicast group for exchanging MDP control messages.

A multicast router in the mobile host's cell initiates a joining to the multicast address assigned to the mobile host. Typically this router coexists with the base station in the cell. As was presented in the description of the DCM approach this join message is propagated to the DCR inside the area that serves the mobile host's multicast address. Then, the DCR sends to the MDP control multicast address a MDP control message informing that now the mobile host has registered . In the IETF proposal the home agent is the only place that knows the mobile host's current position and all communication with the mobile host is done via the home agent. In our approach this is avoided since within each area there is a DCR that is aware of the mobile host.

In order to reduce packet latency and losses during a handoff, advance registration can be performed. The goal is that when a mobile host moves to a new cell, the base station in the new cell has already started receiving data destined to the mobile host. The mobile host continues receiving the data without disruption. There are several ways to perform this:

- A base station that anticipates the arrival of a mobile host initiates joining to the multicast address assigned to the mobile host.
- In the case where a bandwidth is not expensive, all neighbouring base stations can start receiving data destined to a mobile host. This guarantees that there would be no latency and packet losses during a handoff.

When a host wants to send data to a mobile host it sets the destination address to the multicast address assigned to the mobile host and sends the packet as a local multicast. This multicast packet is sent encapsulated to the DCR inside the area that serves the mobile host's multicast address. From this DCR a multicast packet is delivered to the DCR(s) where a mobile host has registered, by using point-to-point tunnels mechanism described before. Upon receiving encapsulated data for the mobile host, the DCRs decapsulate multicast data and forwards the data along established subtrees to base stations. A mobile host receives data only from a base station in its current cell.

The IP multicast architecture with receiver-initiated joins and prunes and soft-state to time out forwarding state in routers on the delivery tree, enables that the advance registration is performed efficiently. At its current cell the mobile host receives data along a distribution subtree that is established for the mobile host's multicast address. This tree is rooted at the DCR and maintained with periodical sending of join messages. Routers on the distribution tree keep forwarding information for a timeout, even if the base station stops sending joins because of handoff. When a base station in the neighbouring cell anticipates arrival of the mobile host, it begins a joining process for the multicast group assigned to the mobile host.This joining is terminated when a join message reaches the router that is already on the distribution tree. When the cells are close to each other, joining is terminated at the lowest branching point in the distribution tree.

In this paper we do not address the problems of using multicast routing to support end-to-end unicast communication. These problems are related to protocols such as: TCP, ICMP, IGMP, ARP. For these issues see [7].

Here we give an initial solution to this problem. We propose to have a special range of unicast addresses that are routed as multicast addresses. In this way, packets destined to the mobile host are routed by using a multicast mechanism. Conversely, at the end systems, these packets are considered as unicast packets and standard unicast mechanisms are applied.

Conclusions

We have considered the problem of multicast routing in a large single domain network with very large number of multicast groups with a small number of receivers. Such a case occurs, for example, when multicast addresses are statically allocated to mobile terminals, as a mechanism to manage Internet host mobility. Our proposal, called Distributed Core Multicast

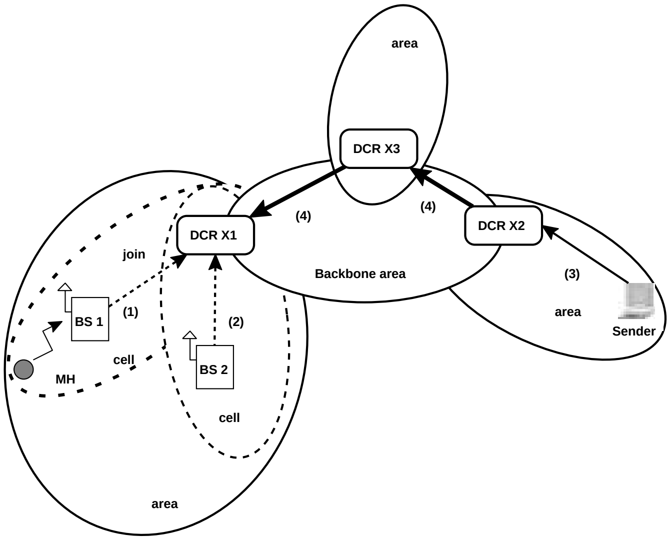

Figure 2: The mobile host (MH) is assigned the multicast address m. Three DCRs, X1, X2 and X3 serve m. Step (1): Base station (BS1) sends a join message for m towards X1. Step (2): The advance registration for m in a neighbouring cell is done by BS2. X1 informs X2 and X3 that it has registered mobile host for m. Step(3): Sender sends a packet to multicast group m. This packet gets delivered through the backbone to X1. Step (4): X1 receives encapsulated multicast data. From X1 data is forwarded to BS1 and BS2. MH receives data from BS1.

(DCM) is based on an extension of the centre-based tree approach. We introduce an architecture based on several core routers, called Distributed Core Routers (DCRs) and a special protocol between them. The objectives we wanted to achieve are: (1) avoid state information in backbone routers, (2) avoid triangular routing across expensive backbone links, (3) scale well with the number of multicast groups. We have also described how our approach can be used to support mobility in the Internet.

References

[1] Network Simulator. Available from http://www-mash.cs.berkeley.edu/ns.

[2] A. Ballardie. Core Based Trees (CBT) Multicast Routing Architecture. RFC 2201, September 1997.

[3] S. Deering and R. Hinden. Internet Protocol, Version 6 (IPv6) Srecification. Technical report, RFC 1883, 1995.

[4] Deborah Estrin, Mark Handley, Ahmed Helmy, Polly Huang, and David Thaler. A Dynamic Mechanism for Rendezvous-based Multicast Routing. ACM/IEEE, May, 1997.

[5] D. Estrin et.all. Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification. RFC 2117, June 1997.

[6] John Ioannidis, Dan Duchamp, and Gerald Q. Maguire Jr. IP-based Protocols for Mobile Internetworking. In *Proc.of SIGCOMM'91*, Zurich, Switzerland, September 1991.

[7] Jayanth Mysore and Vaduvur Bharghavan. A New Multicasting-based Architecture for Internet Host Mobility. In *The Third Annual ACM/IEEE International Conference on Mobile Computing and Networking (Mobicom 97).*

[8] C. Perkins. IP Mobility Support, Network Working Group. RFC 2002, October 1996.

[9] Fumio Teraoka, Yasuhiko Yokote, and Mario Tokoro. A Network Achitecture Providing Host Migration Transparency. In *Proc.of ACM SIGCOMM'91.*

[10] David G. Thaler and Chinya V. Ravishankar. Distributed Center-Location Algorithms. *IEEE JSAC*, 15(3), April 1997.

[11] Vinod Valloppillil and Keith W. Ross. Cache Array Routing Protocol v1.0. Technical report, INTERNET-DRAFT, 1998.

[12] Liming Wei and Deborah Estrin. The Trade-offs of Multicast Trees and Algorithms. In *Proc.of the 1994 International Conference on Computer Communications and Networks*, San Francisco, CA, USA, September 1994.