Distributed Core Multicast (DCM): a multicast routing protocol for many groups with few receivers

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 present a multicast routing protocol called Distributed Core Multicast (DCM). It is intended for use within a large single Internet domain network with a very large number of multicast groups with a small number of receivers. Such a case occurs, for example, when multicast addresses are allocated to mobile hosts, as a mechanism to manage Internet host mobility or in large distributed simulations. For such cases, existing dense or sparse mode multicast routing algorithms do not scale well with the number of multicast groups. DCM is based on an extension of the centre-based tree approach. It uses several core routers, called Distributed Core Routers (DCRs) and a special control protocol among them. DCM aims: (1) avoiding multicast group state information in backbone routers, (2) avoiding triangular routing across expensive backbone links, (3) scaling well with the number of multicast groups. We evaluate the performance of DCM and compare it to an existing sparse mode routing protocol when there is a large number of small multicast groups.

1 Introduction

We describe a multicast routing protocol called Distributed Core Multicast (DCM). DCM is designed to provide low overhead delivery of multicast data in a large single domain network for a very large number of small groups. This 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.

DCM is a sparse mode routing protocol, designed to scale better than the existing multicast routing protocols when there are many multicast groups, but each group has in total a few members.

Recent sparse mode multicast routing protocols, such as the protocol independent multicast (PIM-SM) [4] and the core-based trees (CBT) [2], build a single delivery tree per multicast group that is shared by all senders in the group. This tree is rooted at a single centre router called "core" in CBT, and "rendezvous point" (RP) in PIM-SM.

Both centre-based routing protocols have the following potential shortcomings:

- traffic for the multicast group is concentrated on the links along the shared tree, mainly near the core router;
- finding an optimal centre for a group is a NP-complete problem and requires the knowledge of the whole network topology [12]. Current approaches typically use either an administrative selection of centers or a simple heuristic [10]. Data distribution through a single centre router could cause non optimal distribution of traffic in the case of a bad positioning of the centre router, with respect to senders and receivers. This problem is known as a triangular routing problem.

PIM-SM is not only a centre-based routing protocol, but it also uses source-based trees. With PIM-SM, destinations can start building source-specific trees for sources with a high data rate. This partly addresses the shortcomings mentioned above, however, at the expense of having routers on the source-specific tree keep source-specific state. Keeping the state for each sender is undesirable when the number of senders is large.

Multicast Source Discovery Protocol (MSDP)[5] allows multiple RPs per multicast group in a single share-tree PIM-SM domain. It can also be used to connect several PIM-SM domains together. Members of a group initiate sending of a join message towards the nearest RP. MSDP enables RPs, which have joined members for a multicast group, to learn about active sources to the group. Such RPs trigger a source specific join towards the source. Multicast data arrives at the RP along the source-tree and then is forwarded along the group shared-tree to the group members. [13] proposes to use the MSDP servers to distribute the knowledge of active multicast sources for a group.

DCM is based on an extension of the centre-based tree approach and is designed for the efficient and scalable delivery of multicast data under the assumptions that we mention above (a large number of multicast groups, a few receivers per group and a potentially a large number of senders to a multicast group).

As a first simplifying step, we consider a network model where a large single domain network is configured into areas that are organised in a two-level hierarchy. At the top level is a single backbone area. All other areas are connected via the backbone(see Figure 1). This is similar to what exists with OSPF[7].

The issues addressed by DCM are: (1): to avoid multicast group state information in backbone routers, (2): to avoid triangular routing across expensive backbone links and (3) to scale well with the number of multicast groups.

The following is a short DCM overview and it is illustrated in Figure 1. 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

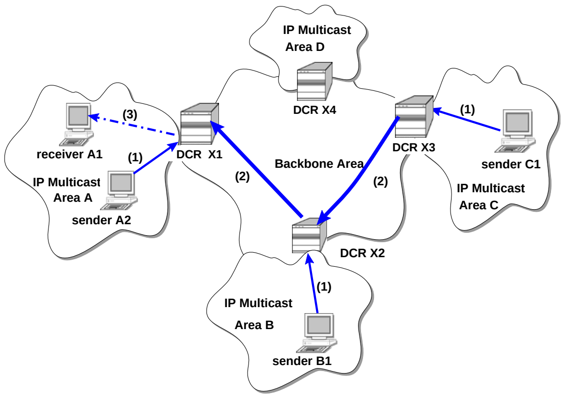


Fig. 1. This is a model of a large single domain network and an overview of data distribution with DCM. In this example there are four non-backbone areas that communicate via the backbone. We show one multicast group M and DCRs X1, X2, X3 and X4 that serve M. Step (1): Senders A2, B1 and C1 send data 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.

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 group. 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 group.

- 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.
- The 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, which does not require that non-DCR backbone routers perform multicast routing.

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 the triangular routing problem common to all centre-based trees, and unlike PIM-SM, is suitable for groups with many sporadic senders. Similar to PIM-SM and CBT, DCM is independent on underlying unicast routing protocol.

In this paper we examine the properties of DCM in a large single domain network. However, DCM is not constrained to a single domain network. Interoperability of DCM with other inter-domain routing protocols is the object of ongoing work.

The structure of this paper is as follows. In the next section we present the architecture of DCM. That is followed by the DCM protocol specification in Section 3. In Section 4 we give a preliminary evaluation of DCM. Section 5 presents how DCM can be used to route packets to the mobile hosts.

2 Architecture of DCM

In this section we describe the general concepts used by DCM. A detailed description follows in Section 3. We group general concepts into three broad categories: (1) hierarchical network model (2) how membership information is distributed and (3) how user data is forwarded.

2.1 Hierarchical Network Model

We consider a network model where a large single domain network is configured into areas that can be viewed as being organised in a two-level hierarchy. At the top level is a single backbone area to which all other areas connect. This is similar to what exists with OSPF[7] and MOSPF[6]. In DCM we use the area concept of OSPF. DCM, unlike MOSPF, does not require link state routing. DCM is independent of the underlying unicast routing protocol.

Our architecture introduces several core routers per multicast group that are called Distributed Core Routers (DCRs). The DCRs are border routers situated at the edge with the backbone. Inside each non-backbone area there can exist several DCRs serving as core routers for the area.

2.2 Distribution of the Membership Information

Regarding the two-level hierarchical network model, we distinguish distribution of the membership information in non-backbone areas and in the backbone area.

Inside non-backbone areas, multicast routers keep group membership information for groups that have members inside the corresponding area. But unlike MOSPF, the group membership information is not flooded inside the area. The state information kept in multicast routers is per group ((*,G) state) and not per source per group (no (S,G) state). If for the multicast group G there are no members inside an area, then no (*,G) state is kept in that area. This is similar to MSDP when it is applied on our network model.

Inside the backbone, non-DCR routers do not keep the membership information for groups that have members in non-backbone areas. This is different from MSDP where backbone routers can keep (S,G) information when they are on the source specific distribution trees from the senders towards RPs. This is also different from MOSPF where all backbone routers have complete knowledge of all areas' group membership. In DCM, the backbone routers may keep group membership information for a small number of reserved multicast groups that are used for control purposes inside the backbone. We say a DCR is labelled with a multicast group when there are members of the group inside its corresponding area. DCRs in different areas run a special control protocol for distribution of the membership information, e.g information of being labelled with the multicast group.

2.3 Multicast Data Distribution

Multicast packets are distributed natively from the local DCR in the area to members inside the area. Multicast packets from senders inside the area are sent towards the local DCR. This can be done by encapsulation or by source routing. This is similar to what exists in MSDP.

DCRs act as packet exploders, and by using the other areas' membership information attempt to send multicast data across the backbone only to those DCRs that need it (that are labelled with the multicast group). DCRs run a special data distribution protocol that try to optimize the use of backbone bandwidth. The distribution trees in the backbone are source-specific, but unlike MSDP do not keep (S,G) information.

3 The DCM Protocol Specification

In this section we give the specification of DCM by describing the protocol mechanisms for every building block in the DCM architecture.

3.1 Hierarchical Network Model Addressing Issues

In each area there are several routers that are configured to act as candidate DCRs. The identities of the candidate DCRs are known to all routers within an area by means of an intra-area bootstrap protocol [3]. This is similar to PIM-SM with the difference that the bootstrap protocol is constrained within an area. This entails a periodic distribution of the set of reachable candidate DCRs to all routers within an area.

Routers use a common hash function to map a multicast group address to one router from the set of candidate DCRs. For a particular group address M, we use the hash function to determine the DCR that serves[1] M.

The used hash function is $h(r(M), DCR_i)$. Function $r(M)$ takes as input a multicast group address and returns the range of the multicast group, while DCR_i is the unicast IP address of the DCR. The target DCR_i is then chosen as the candidate DCR with the highest value of $h(r(M), DCR_j))$ among all j from set $\{1, .., J\}$ where J is the number of candidate DCRs in an area:

$$h(r(M), DCR_i) = \max\{h(r(M), DCR_j), j = 1, .., J\} \tag{1}$$

[1] A DCR is said to serve the multicast group address M when it is dynamically elected among all the candidate DCRs in the area to act as an access point for address M

One possible example of the function that gives the range[2] of the multicast group address M is :

$$r(M) = M\&B \text{ , where } B \text{ is a bit mask.} \tag{2}$$

We do not present here the hash function theory. For more information see [11], [3] and [9]. The benefits of using hashing to map a multicast group to DCR are the following:

- We achieve minimal disruption of groups when there is change in the candidate DCR set. This means that we have to do a small number of re-mappings of multicast groups when there is a change in the candidate DCR set. See [11] for more explanations.
- We apply the hash function $h(.,.)$ as defined by the Highest Random Weight (HRW) [9] algorithm. This function ensures load balancing between candidate DCRs. This is very important, because no single DCR serves more multicast groups than any other DCR inside the same area. We achieve, by this property, that when the number of candidate DCRs increases, the load on each DCR decreases.

All routers in all non-backbone areas should apply the same functions $h(.,.), r(.)$.

Each candidate DCR is aware of all the ranges of multicast addresses for which it is elected to be a DCR in its area. There is one reserved multicast address that corresponds to every range of multicast group address. A DCR joins a reserved multicast address that corresponds to a range of multicast addresses that it serves. This multicast address is used by DCRs in different areas that serve the same range of multicast addresses to exchange control information (see Section 3.3).

3.2 Distribution of membership information inside non-backbone areas

When a host is interested in joining the multicast group M, it issues an IGMP join message.

A multicast router on its LAN, known as the designated router (DR), receives the IGMP join message. The DR determines the DCR inside its area that serves M, as described in the Section 3.1.

The process of establishing the group shared tree is like in PIM-SM [4]. The DR sends a join message towards the determined DCR. 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 M. When a join message reaches the DCR, this DCR becomes labelled with the multicast group M. In this way, the delivery subtree , for the receivers of the multicast group M in an area, is established. The subtree is maintained

[2] A range is the partition of the set of multicast addresses into group of addresses. A range to which a multicast group address belongs to is defined by Equation (2). e.g if the bit mask is (hex) 00000009 we get 4 possible ranges of IPv4 class-D addresses.

by periodically refreshing the state information for M in the routers (like in PIM-SM, this is done by periodically sending join messages).

Like in PIM-SM, when the DR discovers that there are no longer any receivers for M, it sends a prune message towards the nearest DCR to disconnect from the shared distribution tree. Figure 2 shows an example of joining the multicast group.

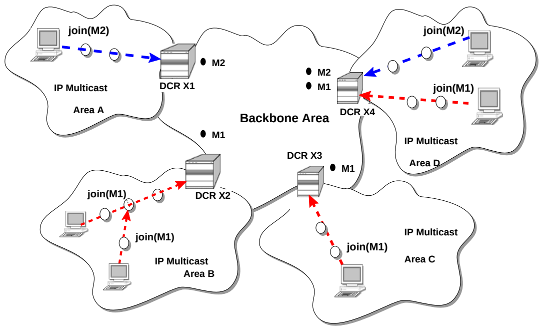


Fig. 2. The figure shows hosts in four areas that join two multicast groups *M1* and *M2*. Four DCRs (X1,X2,X3 and X4) presented in the figure serve the range of multicast addresses where group addresses *M1* and *M2* belong to. A circle on the figure represents multicast routers in non-backbone areas that are involved in the construction of the DCR rooted subtree. These subtrees are showed with dashed lines. X2, X3 and X4 are now labelled with *M1*, while X1 and X4 are labelled with *M2*.

3.3 Distribution of membership information inside the backbone

The Membership Distribution Protocol (MDP) is used by DCRs in different areas to exchange control information. As said in Section 3.1, within each non-backbone area, for each range of multicast addresses (as defined by Equation (2)) there is one DCR serving that range. DCRs in different areas that serve the same range of multicast addresses are members of the same MDP control multicast group. This group is defined by a MDP control multicast address used for exchanging control information. A DCR joins as many MDP control multicast groups as the number of ranges of multicast addresses it serves. There are as many MDP control multicast groups as there are possible ranges of multicast addresses. We do not propose a specific protocol for maintaining the multicast tree for the MDP multicast group. This can be done by means of an existing multicast routing protocol (e.g CBT).

DCRs that are members of the same MDP control multicast group exchange the following control information:

– **periodical keep-alive messages.**

- **unicast distance information.** Each DCR sends, to the corresponding MDP control multicast group, information about the unicast distance from itself to other DCRs that it has learned to serve the same range of multicast addresses. This information comes from existing unicast routing tables and it is used for the distribution of multicast data among the DCRs.
- **multicast group information.** A DCR, which is labelled with the multicast group M, informs DCRs in other areas responsible for M that it has receivers for M. In this way, every DCR keeps a record of every other DCR that has at least one member for a multicast address from the range that the DCR serves. A DCR should notify all other DCRs when it becomes labelled with a new multicast group or no longer labelled with a multicast group.

3.4 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 M. We call this DCR the source DCR. The DR encapsulates the multicast data packet (IP-in-IP) and sends it with a destination address equal to the address of the source DCR. The 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.

3.5 Data distribution in the backbone

The multicast data for the group M is 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 a minimal state information while at the same time backbone bandwidth is used efficiently. We propose a solution that can be applied in the Internet today. It uses point-to-point tunnels to perform data distribution among DCRs. With this solution, non-DCR backbone routers do not keep any state information related to the distribution of the multicast data in the backbone.

Point-to-Point Tunnels The DCR that serve the multicast group M keeps the following information: (1) a set V of DCRs that serve the same range to which M belongs; (2) information about unicast distances between each pair of DCRs from V; (3) the set L of labelled DCRs for M. The DCR obtains this information by exchanging the MDP control messages with DCRs in other areas. In this way, we present the virtual network of DCRs that serve the same range of multicast group addresses by means of an undirected complete graph $G = (V, E)$. V is defined above, while the set of edges E are tunnels between each pair of DCRs in V. Each edge is associated with a cost value that is equal to an inter-DCR unicast distance.

The source DCR, called S, calculates the optimal tree that spans the labelled DCRs. In other words, S finds the subtree $T = (V_T, E_T)$ of G that spans the set of nodes L such that $cost(T) = \sum_{e \in E_T} cost(e)$ is minimised. We recognise this problem as the Steiner tree problem. Instead of finding the exact solution, that is a NP-complete problem, we introduce a simple heuristic called Shortest Tunnel Heuristic (STH). STH consists of two phases. In the first phase a greedy tree is built by adding one by one the nodes that are closest to the tree under construction, and then removing unnecessary nodes. The second phase is further improving the tree established so far.

Phase 1: Build a greedy tree

- **Step 1**: Begin with a subtree T of G consisting of the singe node S. $k = 1$.
- **Step 2**: if $k = n$ then goto **Step 4**. n is the number of nodes in set V.
- **Step 3**: Determine a node $z_{k+1} \in V$, $z_{k+1} \notin T$ closest to T (ties are broken arbitrarily). Add the node z_{k+1} to T. $k = k + 1$. Goto **Step 2**.
- **Step 4**: Remove from T non-labelled DCRs of degree[1] 1 and degree[2] 2 (one at a time).

Phase 2: Improve a greedy tree

STH can be further improved by two additional steps:

- **Step 5**: Determine a minimum spanning tree for the subnetwork of G induced by the nodes in T (after the step 4).
- **Step 6**: Remove from the minimum spanning tree non-labelled DCRs of degree 1 and 2 (one at a time). The resulting tree is the (suboptimal) solution.

Figure 3, Figure 4 and Figure 5 illustrate three examples of the usage of STH. Nodes X1, X2, X3 and X4 present four DCRs that serve the multicast group M. In all examples the source DCR is X1, and the labelled DCRs for M are X2 and X4. For the first two examples, the tree that is obtained by the first phase cannot be further improved by steps 5 and 6. In the third example, steps 5 and 6 give improvements in terms of cost of the resulting tree.

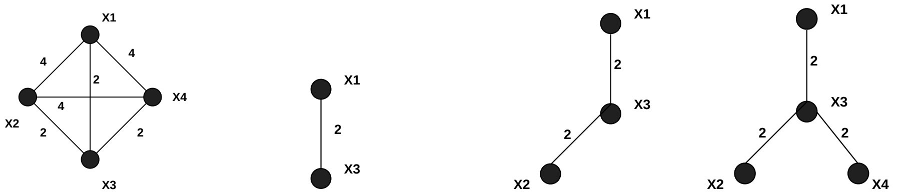
The source DCR applies STH to determine the distribution tunnel tree from itself to the list of labelled DCRs for the multicast group. The source DCR puts inter-DCR distribution information in the form of an explicit distribution list in the end-to-end option field of the packet header. Under the assumption that there is a small number of receivers per multicast group, the number of labelled DCRs for a group is also small. Thus, an explicit distribution list that completely describes the distribution tunnel tree is not expected to be long.

When a DCR receives a packet from another DCR, it reads from the distribution list whether it should make a copy of the multicast data and of the identities of the DCRs where it should send multicast data by tunneling. Labelled DCRs deliver data to local receivers in the corresponding area. An example that shows how multicast data is distributed among DCRs is presented in Figure 6.

---

[1] Degree of a node in a graph is the number of edges incident with a node

[2] A node of degree 2 is removed by its two edges being replaced by a single edge (tunnel) connecting the two nodes adjacent to the node being removed. The source DCR is never removed from a graph



(a) the four node complete graph

(b) start with X1; add X3 to the tree because it is closer to X1 than X2 and X4

(c) X2 is added to the tree

(d) result of STH up to step 4; since X3 is of degree 3 it is not removed from the tree; this is STH solution since the tree cannot be improved by the steps 5 and 6.

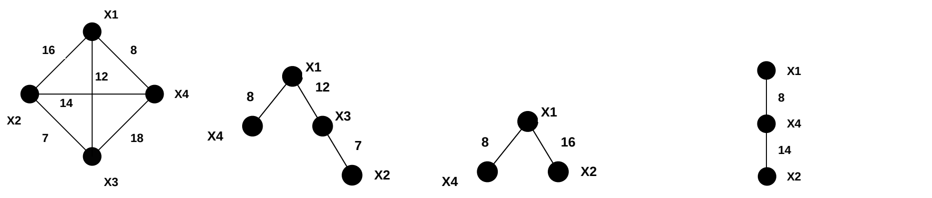
Fig. 3. The first example of the application of STH on the complete graph

(a) the four node complete graph

(b) result of STH up to step 4

(c) X3 is of degree 2 and it is not the labelled DCR; it is removed from the tree; X2 and X4 are connected with the edge (a tunnel between them); this is STH solution since the tree cannot be improved by the steps 5 and 6.

Fig. 4. The second example of the application of STH on the complete graph



(a) the four node complete graph

(b) result of STH up to step 4

(c) X3 is removed from the tree; result of STH up to step 5

(d) Steps 5 and 6: minimum spanning tree of X1, X2 and X4; this is STH solution

Fig. 5. The third example of the application of STH on the complete graph

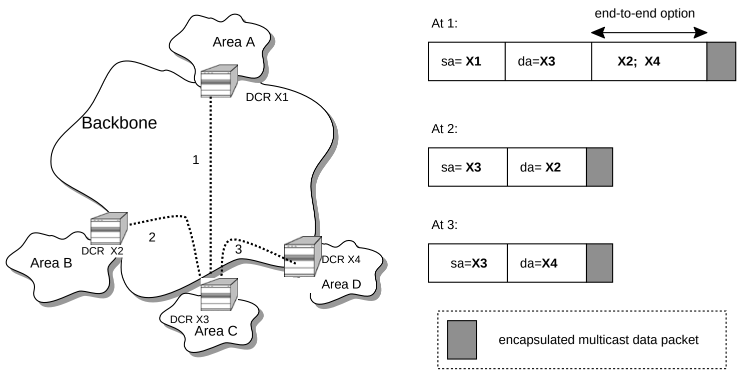


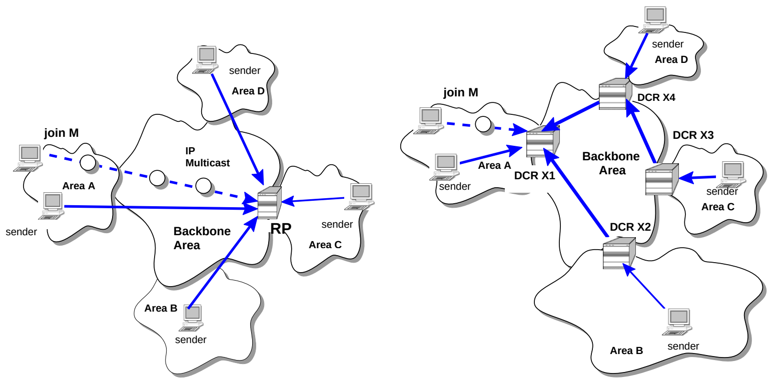
Fig. 6. Figure presents an example of inter-DCR multicast data distribution by using point-to-point tunnels. The source DCR is X1 and labelled DCRs are X2 and X4. X1 calculates the distribution tunnel tree to X2 and X4 by applying STH. Assume that the result of STH gives the distribution tunnel tree consisting of edges X1-X3, X3-X2 and X3-X4. This is similar to the example presented in Figure 3. Then X1 sends the encapsulated multicast data packet to X3. In the end-to-end option field of the packet, a distribution list is contained. X3 sends two copies of multicast data: one to X2 and the other to X4. On this figure are also presented packet formats at various points (points 1, 2 and 3) on the way from X1 to X2 and X4. A tunnel between the two DCRs is shown with the dash line.

3.6 Data distribution inside non-backbone area

A DCR receives encapsulated multicast data packets either from a source that is within its area, or from a DCR in another area. A DCR checks if it is labelled with the multicast group that corresponds to the received packet, i.e whether there are members of the multicast group in its area. If this is the case, a DCR forwards the multicast packet along the distribution subtree that is already established for the multicast group (as is described in Section 3.2).

4 Preliminary Evaluation of DCM

In this section we examine DCM performance under following assumptions: large number of multicast groups, a few receivers per group and a potentially large number of senders to a multicast groups. We show that, under these assumptions, DCM performs better than the PIM-SM shared-tree multicast routing protocol.



(a) PIM-SM Shared-tree (b) DCM

Fig. 7. The figure presents one member of the multicast group M in area A and four senders in areas A, B, C and D. Two different approaches for data distribution are illustrated: the PIM-SM shared-tree case and DCM. In the case of DCM within each area there is one DCR that serves M. In PIM-SM one of the DCRs is chosen to be the centre router (RP) . With PIM-SM, all senders send encapsulated multicast data to the RP. In DCM each sender sends encapsulated multicast data to the DCR inside their area. With PIM-SM, multicast data is distributed from the RP along established distribution tree to the receiver (dashed line). With DCM, data is distributed from source DCRs (X1, X2, X3 and X4) to a receiver by means of point-to-point tunnels (full lines in the backbone) and the established subtree in Area A (dashed line)

We have implemented DCM using the Network Simulator (NS) tool [1]. To examine the performance of DCM in a realistic manner, we performed simulations on a single-domain network model consisting of four areas connected via the backbone area. Figure 7 illustrates the network model used in simulations where areas A,B, C and D are connected via the backbone. The whole network contains 128 nodes. We examined the performance under realistic conditions: the links on the network were configured to run at 1.5Mb/s with a 10ms delay between hops. The link costs in the backbone area are higher than the costs in other areas.

We analyse the following characteristics: size of the routing table, traffic concentration in the network and control traffic overhead.

- **The amount of multicast router state information**
  DCM requires that each multicast router maintains a table of multicast routing information. In our simulations, we want to check the size of multicast router routing table. The routing table size becomes an especially important issue when the number of senders and groups grows, because router speed and memory requirements are impacted.
  We performed a number of simulations. In all the simulations, we use the same network model presented in Figure 7, but with different numbers of multicast groups. For each multicast group there is only one receiver and 20 senders.
  Within each area, there is more than one candidate DCR. The hash function is used by routers within the network to map a multicast group to one DCR in the corresponding area. We randomly distributed membership among a number of active groups. For every multicast group, one receiver in the network is chosen randomly. In the same way, senders are chosen.
  The same scenarios were simulated with PIM-SM applied as the multicast routing protocol. In PIM-SM, candidate RP routers are placed at the same location as candidate DCRs in the DCM simulation.
  We verified that among all routers in the network, routers with the largest routing table size are DRCs in the case of DCM. In the case of PIM-SM those are RPs and backbone routers. We define the most loaded router as the router with the largest routing table size. Figure 8 shows the routing table size in the most loaded router for the two different approaches. Figure 8 illustrates that the size of the routing table of the most loaded DCR is increasing linearly with the number of multicast groups. The most loaded router in PIM-SM is in the backbone. As the number of multicast groups increases, the size of the routing table in the most loaded DCR becomes considerably smaller than the size in the most loaded PIM-SM backbone router.
  As it is expected, routing table size in RPs is larger than in DCRs. This can be explained by the fact that the RP router in the case of PIM-SM is responsible for the receivers and senders in the whole domain, while DCRs are responsible for receivers and senders in the area where the DCR belongs.
  For non-backbone routers, simulation results show that with the placement of RPs at the edge with the backbone there is not a big difference in their routing table sizes for two the approaches. Otherwise, if the location of RPs is elsewhere inside the area, non-backbone routers have smaller routing table

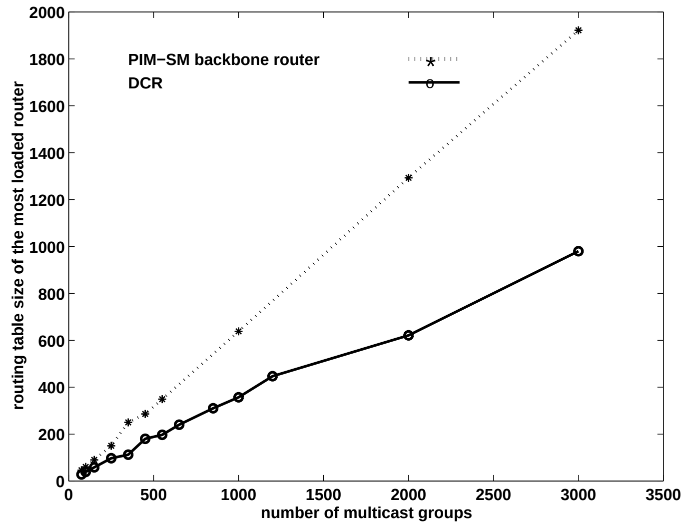


Fig. 8. Routing table size for the most loaded routers

size in the case when DCM is applied as the multicast routing protocol than in the case of PIM-SM.
Figure 9 illustrates the average routing table size in the backbone routers for the two routing protocols. In case of PIM-SM this size is increasing linearly with the number of multicast group. With DCM all join/prune messages from receivers in non-backbone areas are terminated at the corresponding DCRs situated at the edge with the backbone. Thus, in DCM non-DCR backbone routers need not keep multicast group state information for groups with receivers inside non-backbone areas. Backbone routers may keep group membership information only for a small number of MDP control multicast groups.

- **Traffic concentration**
  In the shared-tree case of PIM-SM, every sender to a multicast group sends encapsulated data to the RP router uniquely assigned to that group within the whole domain. This is illustrated in Figure 7(a) where all four senders to a multicast group send data to a single point in the network. This increases traffic concentration on the links leading to the RP.
  With DCM, converging traffic is not sent to a single point in the network because each sender sends data to the DCR assigned to a multicast group within the corresponding area (as presented in Figure 7(b)).
  In DCM, if all senders and all receivers are in the same area, data is not forwarded to the backbone. In that way, backbone routers don't forward the local traffic generated inside an area. Consequently, triangular routing across expensive backbone links is avoided.
- **Control traffic overhead**

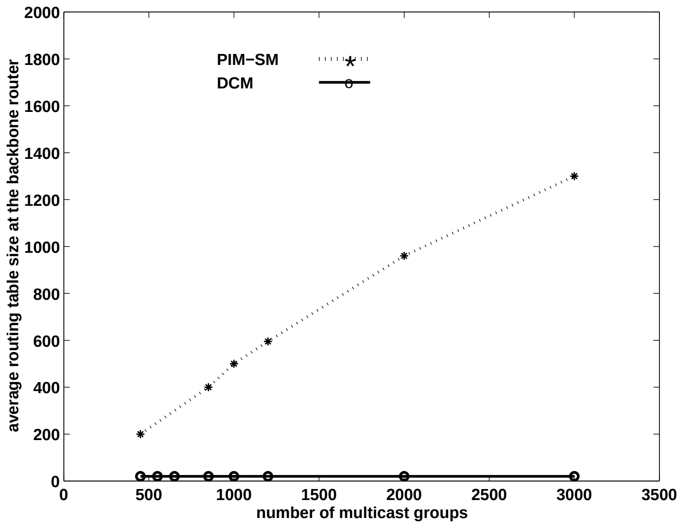


Fig. 9. Average routing table size at the backbone router

Join/prune messages are overhead messages that are used for setting up, maintaining and tearing down the multicast data delivery subtrees. In our simulations we wanted to measure the number of such messages that are exchanged in two cases when DCM and PIM-SM are used as the multicast routing protocols. Simulations have shown that in DCM the number of join/prune messages is 20% smaller than in PIM-SM. This result can be explained by the fact that in DCM all join/prune messages from the receivers in the non-backbone areas are terminated at the corresponding DCRs inside the same area, close to the destinations. In PIM-SM join/prune messages must reach the RP that may be far away from the destinations.

In DCM, DCRs exchange the MDP control messages. The evaluation of the overhead of these messages depends on the group joining/leaving dynamicity and updating frequency. This is left for the future work.

5 Application of DCM in the new mobility management scheme

In this section we show how DCM can be used for a new mobility management approach based on multicasting. When a visiting mobile host arrives into the new domain it is assigned a temporary multicast address. This is the care-of address that the mobile keeps as long it stays in the same domain. This is unlike Mobile IP [8] proposal where the mobile host does a location update after each migration and informs about this its possible distant home agent.

We propose to use DCM as the mechanism to route packets to the mobile hosts. As explained in Section 2.1, 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 packets to the mobile host. As said before, the DCRs run the MDP control protocol and are members of a MDP control multicast group for exchanging MDP control information.

A multicast router in the mobile host's cell initiates a joining the multicast group assigned to the mobile host. Typically this router coexists with the base station in the cell. As described in Section 3.2 the 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 group a MDP control message when the mobile host is registered.

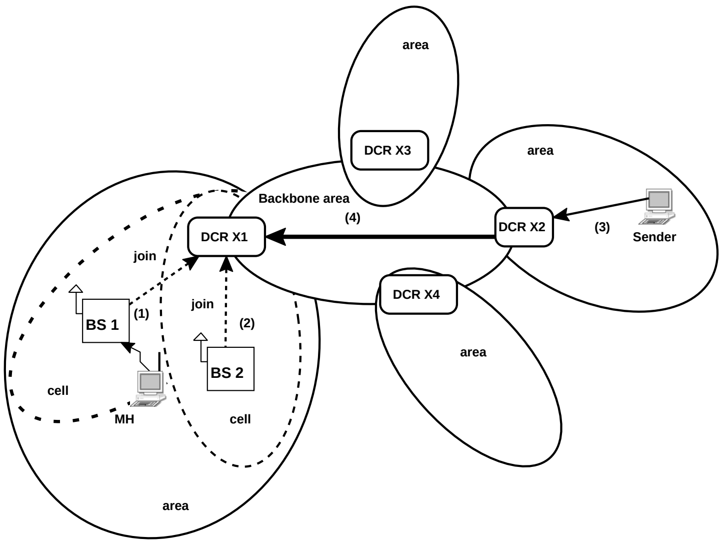


Fig. 10. The mobile host (MH) is assigned multicast address M. Four DCRs, X1, X2, X3 and X4 serve M. Step (1): Base station BS1 sends a join message for M towards X1. X1 informs X2, X3 and X4 that it has a member for M. Step (2): Advance registration for M in a neighbouring cell is done by BS2. Step(3): The sender sends a packet to multicast group M. Step (4): The packet gets delivered through the backbone to X1. Step (5): X1 receives encapsulated multicast data packet. From X1 data is forwarded to BS1 and BS2. MH receives data from BS1.

In order to reduce packet latency and losses during a handover, 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 should already started receiving data for the mobile host. The mobile host continues to receive the data without disruption. There are several ways to perform this:

- A base station that anticipates[1] the arrival of a mobile host initiates joining the multicast address assigned to the mobile host. This is illustrated in one example in Figure 10.
- In the case where a bandwidth is not expensive on the wired network, 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 handover.

A packet for the mobile host reaches all base stations that joined the multicast group assigned to the mobile host. At the same time the mobile host receives data only from a base station in its current cell. A base station that receives a packet on behalf of the mobile host that is not present in its cell can either discard a packet or buffer it for a certain interval of time (e.g. 10ms). Further research is needed to determine what is the best approach.

In this document 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. A simple solution to this problem could be 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.

6 Conclusions

We have considered the problem of multicast routing in a large single domain network with a very large number of multicast groups with a small number of receivers. Our proposal, called Distributed Core Multicast (DCM) is based on an extension of the centre-based tree approach. DCM uses several core routers, called Distributed Core Routers (DCRs) and a special control protocol among them. The objectives achieved with DCM are: (1) avoiding state information in backbone routers, (2) avoiding triangular routing across expensive backbone links, (3) scaling well with the number of multicast groups. Our initial results tend to indicate that DCM performs better than the existing sparse mode routing protocols in terms of multicast forwarding table size. We have presented an example of the application of DCM where it is used to route packets to the mobile hosts.

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.

[1] The mechanism by which the base station anticipates the arrival of the mobile host is out of the scope of this paper

3. Deborah Estrin, Mark Handley, Ahmed Helmy, Polly Huang, and David Thaler. A Dynamic Mechanism for Rendezvous-based Multicast Routing. In *Proc. of IEEE INFOCOM'99*, New York, USA, March 1999.
4. D. Estrin et.al. Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification. RFC 2117, June 1997.
5. D. Farinacci et. al. Multicast Source Discovery Protocol (MSDP). Internet Draft(work in Progress), June 1998.
6. J. Moy. Multicast Extensions to OSPF. RFC 1584, 1994.
7. J. Moy. OSPF version 2. RFC 1583, 1994.
8. C. Perkins. IP Mobility Support, Network Working Group. RFC 2002, October 1996.
9. D. G. Thaler and C. V. Ravishankar. Using Name-Based Mappings to Increase Hit Rates. *IEEE/ACM Transactions on Networking*, 6(1), February 1998.
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. Internet Draft(work in Progress), 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.
13. Li. Yunzhou. Group Specific MSDP Peering. Internet Draft(work in Progress), June 1999.