COMP 535

Multicasting

Simon Foucher (260 223 197)

McGill University

Monday, Nov 9th 2009

Ever since its initial proposition in 1990 by Deering, multicasting has generated lots of research interest and is now built into every IP router and often deployed within enterprise networks. However there is little crossprovider global deployment of multicast because the demand for multicast applications does not justify the complexity of its deployment. Multicasting is the ability for a communication network to accept a message from a single source and to forward it to many specific recipients while minimizing overhead, i.e. by transmitting at most a single copy of the message on any given network link. It can be resumed in 2 actions: joining/leaving a group and sending data to a group. The main issues is that WAN were designed to accommodate point-to-point communication. The need for efficient low cost applications of multicast services is driven by the rise of online gaming, file sharing, RSS, IPTV and other applications. Because of its complexity, most multicast operations are now implemented at the application level.

 WANs can be modeled by a graph where each edge has a weight corresponding to one of its property, here listed in order of priority: low cost, low delay, scalability, support for dynamic multicast groups, survivability to broken links and fairness. Most WANs are symmetric, therefore can be represented by undirected graphs. Routing unicast is essentially solving the graph for shortest path. In multicast, we are interested in the minimum weight tree which spans all the vertices in the multicast group. There are 2 types of multicast communication which yield different optimizations: source specific (data sent only by a single vertex), and group specific (all vertices can communicate with one another).

The multicast optimization problem is called the ‘Steiner tree Problem in Networks’ (SPN). Given an undirected graph G, a cost C for every edge and a list of vertices M, build a tree T which spans M such that the sum of all its edge’s costs is minimized. Some trivial cases include unicast, broadcast and when G is already a tree. Since in communication networks M << G, it may be impossible to find a Steiner tree in a reasonable amount of time, so we usually approximate SPN. These ‘good-quality’ solutions are guaranteed to be at most β times costlier than the optimal solution, and for most of them β = 2.

The Multicast Backbone (MBone) was created in 1992, which is a virtual network sitting on top of the internet to provide multicasting support. The MBone consists of many islands which contain a host running the ‘mrouted’ daemon and supports multicasting within their domains. The islands are connected to one another via P2P IP connections. Multicasting is implemented using many types of protocols which interact with each other to provide full multicast support.

Multicasting on the Internet is implemented with three types of protocols. The first type is employed by a host to join and leave a multicast group, like Internet Group Management Protocol. IGMP is used between a host and the immediately neighboring multicast router. Multicast groups are identified by a class D IP, where 224.0.0.1 is assigned to the permanent group of all IP hosts and is used to address all multicast hosts. Host queries addressed to 224.0.0.1 are periodically sent to discover which groups have members attached on the network. Hosts respond to a query by generating host membership reports reporting each multicast group to which they belong.

The second type of protocol, Multicast Interior Gateway Protocol, enables multicast communication within an AS. The DVRM protocol assigns each link a routing cost used for constructing the RSTP and routers periodically exchange multicast routing updates to build multicast routing tables. This can get very costly over a large AS, so the Core Based Tree protocol was also developed. CBT architecture only requires routers to store routing information for every active group (i.e., per tree) as opposed to storing information for every active (source, group) pair. MOSPF multicasts packets over the shortest-path tree within an AS using the OSPF link database. It extends OSPF by adding a group membership LSA. Routers use IGMP to keep track of group membership information on its attached network, and flood the group membership LSA within the AS. When a router receives a multicast packet, it computes a shortest-path tree and forwards the packet accordingly. Limitations of protocols that impose a single shared tree per group are overhead of broadcasting packets when group members sparsely populate the internet and supporting good-quality distribution. Protocol Independent Multicast was designed to address this and has two modes of operation: PIM Dense Mode (PIM-DM), which employs an RSPT (similar to DVMRP), and PIM Sparse Mode (PIM-SM), which employs unidirectional shared trees. PIM-DM multicast tree construction is data-driven, and is used only if data rates exceed a certain value. PIM-SM employs per-group rendezvous points which are used by senders to announce their existence, and by receivers to learn about new senders of a group.

The third type of protocol is employed by border routers to allow multicast communication across ASes. The Border Gateway Multicast Protocol consists of two components: MIGP and BGMP. The border router employs the MIGP component to participate in the MIGP protocol within the AS, and the BGMP component to construct a bidirectional center-based tree with other border routers. In BGMP, the root of the center- based tree is an entire AS rather than a single router. The root AS of a multicast address is the AS which has claimed the multicast address. Border routers set up a TCP connection between themselves, and exchange BGMP messages. When group memberships change, border routers send incremental join/prune updates to one another.

A simpler approach to implementing IP multicast is called Free Riding Multicast. FRM has 2 major advantages: it eliminates the need for multicast route computation by ‘Free Riding’ existing unicast routes and offers network operators a familiar framework by using the same channels as BGP.

FRM decouples membership discovery from route discovery. This offers the advantage that once group members are known; any source can construct the multicast tree from its unicast routes to each members. FRM can be viewed as extending MOSPF to the inter-domains, which is difficult because of incomplete network map at the inter-domain level.

In terms of protocol FRM, requires extending BGP to carry group membership information and that AS filter some group for a downstream customer AS. Because group membership is explicitly advertised through BGP, an ISP has ultimate control over which groups its customers subscribe to because it can simply drop undesired groups from its BGP advertisement. FRM avoids the contentious selection of RPs and new inter-domain protocols by piggybacking membership state over BGP. It also offers centralized route construction because the multicast tree is computed by the source’s border router using existing unicast routes. FRM incurs the overhead of advertising and maintaining group membership. FRM disseminates membership information more widely than traditional protocols and hence incurs greater overhead. Finding the next hops requires that the access border router scan its entire BGP table and that intermediate nodes decipher the encoded tree. FRM faces the challenge of achieving this in a manner that is both scalable and fast.

FRM comprises two separable components: group membership discovery and multicast packet forwarding. Advertising group membership is done by augmenting border router’s BGP advertisements with a description of the group addresses active within its domain. Forwarding is done differently at the border router in the access domain for the source, and border routers in the transit core. In the first case, caching replaces the linear scan of the BGP RIB’s GRP BF state by an exact-match lookup on cached forwarding state and, if needed, should be mostly achievable in line cards. In the second case, FRM uses a “shim” header above the IP header which encodes the tree edge from autonomous system to another.

FRM represents a different approach to implementing multicast. It is simpler in the wide area (no distributed tree construction), easier to configure (no need to place RPs), and allows providers to work within the familiar BGP framework to handle inter-provider issues. These features come at a cost of reduced efficiency and greater demands on border routers; a tradeoff that we believe is worth exploring given technology trends. FRM tackles a purely technical barrier to deployment and other barriers do exist. However, given the growing adoption of Internet broadcasting, massively multiplayer games, and other networked applications we conjecture the time may be right to revisit IP multicast and re-evaluate its chances.