MULTICAST ROUTING PROTOCOLS IN A MULTIPLE DOMAIN NETWORK

John McCann ¹, Uwe Wagner ¹, Miro Stravs ²
¹ Juniper Networks, London
² Smart Com, Ljubljana

I. INTRODUCTION
Multicast services are based on the deployment of specific routing protocols in the network. Several protocols are required to enable IP multicast over multiple domains. Inter-domain routing is one of the key points concerning multicast routing. Several architectures are proposed in IETF specification groups, using different protocols (MSDP and MBGP,...). Concerning intra-domain routing, PIM and IGMP protocols are de facto standards. These protocols have been developed independently, and each has its own set of specific terms. The multiplicity of protocol development makes it difficult to create a workable, integrated multicast implementation. The intention of this paper is to provide an illustration of an inter-domain multicast routing system, end to end across multiple domains.

Figure 1 shows a simple network, and will serve as a reference for subsequent discussion in this paper. The diagram shows two interconnected autonomous systems (AS) with fully functional unicast routing. We will show how to enable these two autonomous systems to route multicast traffic between them. Each step will concentrate on a part of this diagram and describe the mechanisms that need to be put in place for operational multicast routing. Specifically, the focus will be on allowing Host-B to receive multicast traffic from Server-A.

II. ROUTING MULTICAST TRAFFIC WITHIN A SINGLE DOMAIN USING PIM-SM
The first step is to get multicast routing up and running within a single domain. Protocol Independent Multicast–Sparse Mode (PIM–SM or PIM) is the multicast routing protocol most suited to this task. We define a ‘domain,’ i.e. a domain in the context of PIM-SM. Multicast routing decisions are based on the source and not on the destination (like unicast). Each router looks at the source of the traffic and determines the closest interface to it (RPF check) and performs RPF checks to obtain the best path to the source. Multicast routing protocols can be classified in dense mode and sparse mode.

Protocol Independent Multicast is an IP multicast routing protocol that has been designed to cope with multicast groups featuring both sparse and dense types. Dense mode uses a flood and prune mechanisms to build a multicast distribution tree rooted at the source (a source-based tree), giving the shortest and most efficient path from source to receiver. However, the use of flooding across the Internet will not scale.

Sparse mode protocols implement a shared distribution tree, rooted at a core router in the network called a rendezvous point (RP). RP entity will keep information of all active sources in a domain. Only network segments that have active receivers, which have explicitly requested the data, will be forwarded the traffic. If a router is connected to a host that wants to receive data, it need to use RPFs to obtain the shortest path to the RP and all the receivers join the tree at the RP. Thanks to the involvement of this third party (the RP) in the communication, senders and receivers are able to communicate without the need of knowing each other. According to this feature, the PIM protocol is a new approach compared to the inefficient flood-and-prune algorithm. Another major advantage of PIM-SM consists in its ability to switch from an RP-centered shared tree (called the RPT – Rendezvous Point Tree) to a source-centered tree (called the SPT – Shortest Path Tree).
Figure 1 shows two separate domains. PIM domains can usually be mapped to BGP autonomous systems, which are generally defined in BGP as a portion of the Internet that is owned and operated by a specific organization. By identifying a separate autonomous system for each organization, we enable the network architects of those organizations to enforce their own routing policies on their own part of the Internet.

The reason for having separate PIM domains for different organizations is slightly different: Mainly, we do not want routing of our own network’s local traffic to depend on the availability of equipment in someone else’s network.

PIM-SM requires the following three major phases to deliver multicast packets from a source to a receiver:

- Build the RPT that delivers packets from the RP to interested listeners
- Build the distribution tree that delivers packets from the source to the RP
- Build the SPT that delivers packets directly from the source to the interested listeners

These phases occur for each source-receiver pair, and the distribution trees for different sources and groups may have reached different phases at any given point of time, depending on the existence of a source and interested listeners for that group.

The order of operation is not strict. Multicast sources can be created before receivers are created, and PIM still enables delivery of multicast packets to any newly appearing receivers.

Traditional IP multicast model, supporting many-to-many communication, is very difficult to realize over Wide Area Networks especially in an inter-domain environment. Too many issues are putting off the commercial deployment of IP multicast based applications and services.

The consideration that today most of the existing multicast applications are based on one-to-many communication convinced part of the Internet multicast community (ISPs in particular) to converge towards single-source multicasting as a simplification of the multicast paradigm and, above all, an immediately deployable inter-domain solution.

It is important to notice that the SSM (Source Specific Multicast) model represents a significant rethinking about the separation of functionality between the Routing Layer and the Application/Control Layer for wide-area multicasting. Specifically, it aims to restrict the role of a multicast routing protocol to building a multicast tree, and forwarding the data packets along that tree. The responsibility of discovering the identity of a multicast source (or equivalently, a multicast service) is left to the application/control layer.

The host-router multicast protocol is IGMPv3 which include the possibility for a host to join to a multicast group identified by the pair (Source, Group). IGMPv3 supports “source filtering” i.e., the ability of an end-system to express interest in receiving data packets sent only by specific sources or from all but some specific sources. This functionality of IGMPv3 is a superset of the capabilities required to realize the SSM model.
III. ROUTING MULTICAST TRAFFIC ACROSS MULTIPLE DOMAINS WITH MSDP

The PIM protocol in itself does not have a mechanism to enable multicast packets from a source in one PIM domain to reach a receiver in another domain. The delivery of multicast packets from the source to the RP in the source’s domain is disconnected from the RPT in the receiver’s domain. Thus, the following two requirements exist for transporting multicast traffic across multiple PIM domains:

- The RP in domains that have receivers must have knowledge of the IP address of active sources
- All routers along the path from the source to the receivers must have a route to the source’s IP address in their RPF table

The first requirement is accomplished by using Multicast Source Discovery Protocol (MSDP). MSDP provides a way to connect multiple PIM-SM domains without one domain depending on the availability of the other. Each domain relies on its own routers to serve as RP.

A RP within a PIM domain will have a MSDP peering relationship with a RP in another domain, which will be established over a TCP connection (well-known port 639) between the two peers. The purpose of this mechanism is to allow domains discover multicast sources from other domains. Each domain will have one or more connections. If the multicast sources are of interest to a domain that has receivers, multicast data is delivered over the normal, source-tree building mechanism in PIM-SM.

MSDP sessions are formed between the RPs of various domains. MSDP-speaking RPs send MSDP SOURCE-ACTIVE (SA) messages to notify the RPs in other domains of active sources. When an RP receives an SA message for a group for which interested receivers exist, the RP delivers the encapsulated data down the RPT to all the receivers in its domain. When the PIM DR receives the multicast packets down the RPT, it will join the SPT directly to the source.

MSDP is not limited to the deployment across different (BGP) domains. It can be used within an autonomous system when it is desired to deploy multiple RPs for the same group ranges.

The second requirement is usually not a concern, because most networks have any-to-any connectivity for unicast traffic, even for addresses in other autonomous systems. The multicast RPF table need not be the same routing table used for unicast routing. In this case the dedicated multicast RPF table must have routes for all potential multicast sources. MBGP is used to populate such an RPF table.

By taking this approach, unicast and multicast traffic will follow the same path, but in opposing directions. For example, a multicast packet traveling from Server-A to Host-B would traverse all the same routers and links, but in the exact opposite order, as a unicast packet traveling from Host-B to Server-A.

Some situations make such congruent routing of unicast and multicast traffic less than optimal. In such situations a table other than the one used for unicast forwarding must be used for multicast RPF. The question is how to populate such a table: How are unicast routes introduced into a separate RPF table, with next-hop information different from the table used for unicast forwarding?

One solution is to configure static routes specifically for the RPF table. Note that static routing for multicast RPF faces the same scalability limitations as static routing for unicast forwarding. Those limitations being: lack of dynamic failover and maintenance burden because changes to topology are not automatically updated.

In real networks, it is desirable to update the entries in the RPF table dynamically. The RPF table consists of unicast routes, so there is no need to invent a new routing protocol. Instead the need is to somehow differentiate between route-control information intended to be used in the unicast-forwarding routing table and the multicast RPF table.

Theoretically this differentiation could be implemented by modifying any of the existing unicast routing protocols. The structure of some protocols, however, facilitates expanded functionality. From the perspective of software developers, BGP is one of the easiest protocols to which to add such functionality. Within the BGP protocol, a capabilities negotiation occurs between peers when they first establish a session. Currently BGP is the only dynamic routing protocol that can differentiate between multiple types of routing information. This capability is designated Multiprotocol Extensions for BGP (MBGP) and is defined in RFC2283. MBGP works identically to BGP in all respects; it simply adds functionality to BGP. When using MBGP for updating dedicated multicast RPF tables, two sets of routes are exchanged in the MBGP updates:

- IPv4 unicast routes
- IPv4 multicast RPF routes

Each set will most likely have duplicated prefixes, but the path information for the same prefix in each set can be different. Not only can multicast RPF routes have different BGP next-hops and therefore potentially different recursive next-hopsthey can also have different information in any of the BGP path attributes.

A further requirement of multicast inter-domain routing was the ability to support incongruent unicast and multicast routing topologies. It’s necessary to be able to create different routing policies for unicast and multicast. MBGP (Multiprotocol Border Gateway Protocol) creates extensions to the widely-used BGP (Border Gateway Protocol) to support this requirement. MBGP adds a multicast-only reachability table to the existing unicast reachability table of BGP. With MBGP, a router can effectively have two BGP...
tables, one for multicast and one for unicast. MBGP establish a tree of domains or a routing hierarchy.

V. CONCLUSION

We came to the conclusion that IP Multicast is already a stable technology which can be introduced as standard service to any major IP backbone now. Especially streaming services are invading the existing IP networks. Solutions for reliable multicast are available, but the standardisation process in this area has not reached the final stage. As different equipment vendors are implementing multicast protocols on different scale it is important in a practical network solution to chose the right vendor equipment. Source Specific Multicast (SSM) seems to be a solution for intra-domain for most of the multimedia IP services planned by the service or network providers. For inter-domain the MSDP combined with MBGP are the optimal solution. Further investigation and tests of the upcoming implementations are needed. A general solution for QoS in an IP Multicast environment is still missing and it is of high interest of the solution implementer to closely follow the rapid developments in this area of IP Multicast.

Abstract:

Inter-domain routing is one of the key points concerning multicast routing. Several architectures are proposed in IETF specification groups. The intention of this paper is to provide an illustration of an inter-domain multicast routing system, end to end across multiple domains. Protocol Independent Multicast–Sparse Mode (PIM–SM or PIM) is the routing protocol most suited to get multicast routing up and running within a single domain. MSDP provides a way to connect multiple PIM-SM domains without one domain depending on the availability of the other. It’s necessary to be able to create different routing policies for unicast and multicast. MBGP (Multiprotocol Border Gateway Protocol) creates extensions to the widely-used BGP (Border Gateway Protocol) to support this requirement. The presented IP Multicast scenario is a working solution which can be introduced as standard service to any major IP backbone.

Literature:

[4] Internal Juniper material