MBONE Deployment Working Group
Internet Draft

David Meyer
University of Oregon

Expiration Date: December 1997 June 1997

Some Issues for an Inter-domain Multicast Routing Protocol

draft-ietf-mboned-imrp-some-issues-02.txt

1. Status of this Memo

2. Introduction

The IETF's Inter-Domain Multicast Routing (IDMR) working group has produced several multicast routing protocols, including Core Based Trees [CBT] and Protocol Independent Multicasting [PIMARCH]. In addition, the IDMR WG has formalized the specification of the Distance Vector Multicast Routing Protocol [DVMRP]. Various specifications for protocol inter-operation have also been produced (see, for example, [THALER96] and [PIMMBR]). However, none of these protocols seems ideally suited to the inter-domain routing case; that is, while these protocols are appropriate for the intra-domain routing environment, they break down in various ways when applied in to the multi-provider inter-domain case.

This document considers some of the scaling, stability and policy issues that are of primary importance in a inter-domain, multi-provider multicast environment.

3. Forwarding State Requirements

Any scalable protocol will have to minimize forwarding state requirements. In the case of dense mode protocols [DVMRP,PIM-DM], routers may carry forwarding or prune state for every (S,G) pair in the Internet. This is true even for routers that may not be on any delivery tree. It seems likely that as multicast deployment scales to the size of the Internet, maintenance of (S,G) state will become intractable.

Shared tree protocols, on the other hand, have the advantage of maintaining a single (*,G) entry for a group's receivers (thus relaxing the requirement of maintaining (S,G) for the entire Internet). However, this is not without its own disadvantages; see the section on "Third-party Resource Dependencies" below.

4. Forwarding State Distribution

The objective of a multicast forwarding state distribution mechanism is to ensure that multicast traffic is efficiently distributed to those parts of the topology where there are receivers. Dense and sparse mode protocols will accept differing overheads based on design tradeoffs. In the dense mode case, the data-driven nature state distribution has disadvantage that data is periodically distributed to branches of the distribution tree which don't have receivers ("Broadcast and Prune" behavior). It seems unlikely that this mechanism will be scalable to Internet-wide case.

On the other hand, sparse mode protocols use receiver-initiated, explicit joins to establish a forwarding path along a shared distribution tree. While the on-demand nature of sparse mode protocols have favorable properties with respect to distribution of forwarding state, it also has the possible disadvantage of creating dependencies on shared resources (again, see the section on "Third-Party Resource Dependencies" below).

5. Forwarding State Maintenance

The many current multicast protocols attempt to accurately and rapidly maintain distribution trees that are as close to the tree of shortest-path routes (as defined by unicast) as possible. This means that the shape of a distribution tree can be rapidly changing. In addition, since distribution trees can be global, they can be subject to high frequency control traffic.

In contrast, the focus in the inter-domain unicast routing environment is on minimizing routing traffic (see, for example, [VILLAM95]), and controlling stability [LABOV97]. The implication is that protocol overhead and stability must be controlled if we hope multicast to scale to Internet sizes. Thus it seems likely that Inter-domain multicast routing protocols will have to do less forwarding state maintenance, and hence be less aggressive in reshaping distribution trees. Note that this reshaping is related to what has been termed "routing flux" (again, see [LABOV97]), since the routing traffic does not directly affect path selection. Rather, the primary effect is to require significant processing resources in a border router. Finally, note that unlike the unicast case, we do not have good data characterizing this effect for multicast routers.

5.1. Data Driven Forwarding State Creation

Another issue with broadcast and prune protocols is that forwarding state is created in a data-driven manner.

5.2. Bursty Source Problem

When a source's inter-burst period is longer than the router state timeout period, some or all of a source's packets can be lost. This effect has been termed the "Bursty Source Problem" [ESTRIN97]. The current set of multicast routing protocols attempt, where possible, to avoid this problem (i.e., maximize response to bursty sources).

6. Mixed Control

Mixing control of topology discovery and distribution tree construction can lead to efficiencies but also imposes various constraints on topology discovery mechanisms. For example, DVMRP [DVMRP] uses topology discovery facilities ("split horizon with poison reverse") to eliminate duplicate packets on a LAN, and to detect non-leaf networks (an upstream router uses this information when pruning downstream interfaces).

On the other hand, PIM [PIM-DM] does not use any topology discovery algorithm features when building delivery trees. However, this independence is not without cost: PIM-DM accepts some duplicates on multi-access LANs as a tradeoff for reduced protocol complexity.

7. Neighbor Model

The current inter-domain unicast routing model has some key differences with proposed inter-domain multicast routing models with respect to neighbor (peer) discovery. In particular, the current set of multicast protocols depend heavily on dynamic neighbor discovery. This is analogous to the situation with intra-domain unicast routing, but is unlike current inter-domain unicast routing, where neighbors are typically statically configured.

The static neighbor configuration model has several benefits for inter-domain routing. First, neighbors are predefined, which is a policy requirement in most cases. In addition, the set of peers in the inter-domain unicast routing system defines the set of possible inter-domain topologies (with the current active topology represented by the collection of AS paths).

Another important difference relates to how inter-domain regions are modeled. For purposes of this document, consider an inter-domain region defined to be a part of an arbitrary topology in which a higher level (inter-domain) routing protocol is used to calculate paths between regions. In addition, each pair of adjacent regions is connected by one or more multicast border routers. Current IDMR proposals (e.g., [HDVMRP], [THALER96]) model an inter-domain region as a routing domain. That is, border routers internetwork between one or more intra-domain regions and an inter-domain region (again, possibly more than one). In this model, inter-networking occurs "inside" router. However, the inter-provider unicast routing model in use today is quite different. In particular, the "peering" between two providers occurs in neither of the provider's routing domains, nor does it occur in some shared "inter-domain" routing domain. The separation provides the administrative and policy control that is

required in today's Internet.

8. Unicast Topology Dependency

Ideally, unicast and multicast topologies are congruent in the Internet. However, since it is frequently difficult to field new facilities (such as IP multicast) in the "core" the Internet infrastructure, there will continue to be many cases in which unicast and multicast topologies are not congruent (either because a region is not multicast capable at all, or because the region is not natively forwarding multicast traffic). Thus, it is unlikely that the entire IPv4 Internet will be able to carry native multicast traffic in the foreseeable future. In addition, various policy requirements will in certain cases cause to topologies to further diverge. The implication is that a successful IDMR will need a topology discover mechanism, or have other mechanisms for dealing with those cases in which unicast and multicast topologies are not congruent.

8.1. Multicast Policies and Unicast Routes

Multicast and unicast packet forwarding algorithms assign different semantics to a unicast route. In particular, if a router B accepts a route from router A covering prefix P, then B will to forward packets "to" those destinations covered by P, using A as the next hop when forwarding unicast packets. However, in the multicast case, the same route means B will accept packets "from" sources covered by P (though not necessarily from A, but through the same interface as is used to reach A). It is this difference in unicast route semantics that makes formulation of precise multicast policy difficult.

9. Third-Party Resource Dependencies

Shared tree protocols require one or more globally shared Rendezvous Points (RPs) [PIM-SM] or Cores [CBT]. The RP or Core effectively serves as the root of a group specific shared tree. Data is sent to the RP/Core for delivery on the shared tree. This means that some groups may have an RP (or core) that is fielded by a third party. For example, if providers A, B and C share a PIM-SM inter-domain region, then there may exist an RP that is mapped to C's multicast border router. In this case, C is hosting a kind of "transit RP" for A and B (A and B register to C to communicate between themselves, even if C has no receivers for the group(s) served by the RP.

10. Traffic Concentration Problem

Traffic can be "concentrated" on a shared tree. This can lead to increased latency or packet loss. However, this is less of a problem in the shared-media exchange point environment.

11. Distant RP/Core Problem

In the shared tree model, if the RP or Core is distant (topologically), then joins will travel to the distant RP/Core, even if the data is being delivered locally. Note that this problem will be exacerbated if the RP/Core space is global; if a router is registering to a RP/Core that is not in the local domain (say, fielded by the site's direct provider), then the routing domain is flat.

12. Multicast Internal Gateway Protocol (MIGP) Independence

A shared tree, explicit join protocol inter-domain routing protocol may require modification to a leaf domain's internal multicast routing mechanism. The problem arises when a domain is running a "broadcast and prune" protocol such as DVMRP or PIM-DM internally while participating in a shared tree inter-domain protocol. In this case, there can be areas of the (internal) topology that has receivers but will not receive inter-domain traffic.

[THALER96] describes interoperability rules to alleviate this problem. Consider the case where a border router has interfaces in an inter-domain shared tree region and a DVMRP region. The rules covering this case state that either the DVMRP region must implement Region Wide Reports [HDVMRP], or the DVMRP component of the border router must be a wildcard receiver for externally reached sources, while the shared tree component is a wildcard receiver for internally reached sources. Alternatively, many current implementations use a "receiver-is-sender" approach (which depends for the most part on RTCP reports) to get around this problem.

13. Encapsulations

Encapsulations should be minimized where ever possible. PIM-SM encapsulates packets sent to the shared tree in PIM Register messages (data can be delivered natively if the last hop router or the RP switches to the shortest path tree). The design of an shared tree inter-domain protocol should avoid the "O(N) Encapsulation" problem: For paths that traverse N administrative domains, the number of encapsulations can approach O(N).

14. Policy Provisions

Current inter-domain unicast routing protocols have a rich and well developed policy model. In contrast, multicast routing protocols have little or no provision for implementing routing policy (administrative scoping is one major exception). A concrete example of this need is the various problems with inadvertent injection of unicast routing tables into the MBONE, coupled with our inability to propagate the resultant large DVMRP routing tables, point out the need for such policy oriented controls.

A simple example illustrates why a successful inter-domain multicast routing protocol will need to have a well developed policy model: Consider three providers, A, B, and C, that have connections to a shared-media exchange point. Assume that connectivity is non-transitive due to some policy (the common case, since bi-lateral agreements are a very common form of peering agreement). That is, A and B are peers, B and C are peers, but A and C are not peers. Now, consider a source S covered by a prefix P, where P belongs to a customer of A (i.e., P is advertised by A). Now, multicast packets forwarded by A's border router will be correctly accepted by B's border router, since it sees the RPF interface for P to be the shared-media of the exchange. Likewise, C's border router will reject the packets forwarded by A's border router because, by definition, C's border router does not have A's routes through its interface on the exchange (so packets sourced "inside" A fail the RPF check in C's border router).

In the example above, RPF is a powerful enough mechanism to inform C that it should not accept packets sourced in P from A over the exchange. However, consider the common case in which P is multi-homed to both A and B. C now sees a route for P from B though its interface on the exchange. Without some form of multi-provider cooperation and/or packet filtering (or a more sophisticated RPF mechanism), C could accept multicast packets sourced by S from A across the shared media exchange, even though A and C have not entered into any agreement on the exchange. The situation described

above is caused by the overloading of the semantics of unicast route (as described above), and the reliance on the RPF check as a policy mechanism.

14.1. "Wrong" RPF Neighbor

The example above illustrates a the problem that, in most current implementations, the RPF check considers only the incoming interface, and not the upstream neighbor (RPF neighbor). This can result in accepting packets from the "wrong" RPF neighbor (the neighbor is "wrong" since, while the RPF check succeeds and the packet is forwarded, the unicast policy would not have forwarded the packet).

14.2. RPF as a Policy Mechanism

In the example above, C is relying on its RPF check to protect it from A's packets. However, not only is RPF too weak enough to cover those cases in which a source prefix is multi-homed (as described in the example above), it is essentially a packet filter and as such is not an attractive policy mechanism.

15. Today's MBONE

Another way to view the policy issues described above is to consider the perspective of unicast reachability. Today's MBONE is comprised of a single flat AS. Further, this AS running a simple distance vector topology discovery protocol. This arrangement is unlikely to scale gracefully or provide the same rich policy control that we find in the unicast Internet. There are additional problems with a flat AS model: the flat AS model fits neither the operational or organizational models commonly found in Internet today.

16. Equal Cost Multipath

A common way to incrementally scale available bandwidth is to provide parallel equal cost paths. It would be an advantage if a multicast routing protocol could support this. However, this would seem difficult to achieve when using Reverse Path Forwarding, so it is unclear whether this goal is achievable.

17. Conclusion

Deployment of a general purpose IP multicast infrastructure for the Internet has been slowed by various factors. One of the primary reasons, however, is the lack of a true inter-domain Multicast Routing Protocol. Several proposals have been advanced to solve this problem, including PIM-SM [PIM-SM], DVMRP [PIMMBR], and Hierarchical DVMRP [HDVMRP]. However, the concerns outlined above have prevented any of these protocols from being adopted as the standard inter-domain multicast routing protocol. Finally, it is worth noting that DVMRP, since it is the common denominator among router vendor offerings, is currently the de-facto inter-domain routing protocol.

18. Security Considerations

Historically, routing protocols used within the Internet have lacked strong authentication mechanisms [RFC1704]. In the late 1980s, analysis revealed that there were a number of security problems in Internet routing protocols then in use [BELLOVIN89]. During the early 1990s it became clear that adversaries were selectively attacking various intra-domain and inter-domain routing protocols (e.g. via TCP session stealing of BGP sessions) [CERTCA9501, RFC1636]. More recently, cryptographic authentication mechanisms have been developed for RIPv2, OSPF, and the proprietary EIGRP routing protocols. BGP protection, in the form of a Keyed MD5 option for TCP, has also become widely deployed.

At present, most multicast routing protocols lack strong cryptographic protection. One possible approach to this is to incorporate a strong cryptographic protection mechanism (e.g. Keyed HMAC MD5 [RFC2104]) within the routing protocol itself. Alternately, the routing protocol could be designed and specified to use the IP Authentication Header (AH) [RFC1825, RFC1826, RFC2085] to provide cryptographic authentication.

Because the intent of any routing protocol is to propagate routing information to other parties, confidentiality is not generally required in routing protocols. In those few cases where local security policy might require confidentiality, the use of the IP Encapsulating Security Payload (ESP) [RFC1825, RFC1827] is recommended.

Scalable dynamic multicast key management is an active research area at this time. Candidate technologies for scalable dynamic multicast key management include CBT-based key management [RFC1949] and the Group Key Management Protocol (GKMP) [GKMPID]. The IETF IP Security Working Group is actively working on GKMP extensions to the

standards-track ISAKMP key management protocol being developed in the same working group.

19. References

[BELLOVIN89] S. Bellovin, "Security Problems in the TCP/IP Protocol Suite", ACM Computer Communications Review, Volume 19, Number 2, pp. 32-48, April 1989.

[CBT] A. Ballardie, et. al., "Core Based Trees (CBT) Multicast -- Protocol Specification --", draft-ietf-idmr-cbt-spec-06.txt, September, 1996.

[CERTCA9501] CERT, "IP Spoofing Attacks and Hijacked Terminal Connections", ftp://ftp.cert.org/cert_advisories/, January 1995.

[DVMRP] T. Pusateri, "Distance Vector Multicast Routing Protocol", draft-ietf-idmr-dvmrp-v3-03, September, 1996.

[GKMPID] H. Harney, "Group Key Management Protocol (GKMP)", draft-harney-gkmp-arch-01.txt && draft-harney-gkmp-spec-01.txt, August 1996, Informational RFC publication is pending.

[HDVMRP] A. Thyagarajan and Steve Deering, "Hierarchical Distance-Vector Multicast Routing for the MBone", In Proceedings of the ACM SIGCOMM, pages 60-66, October, 1995.

[LABOV97] C. Labovitz, et. al., "Internet Routing Instability", Submitted to SIGCOMM97.

[PIMARCH] D. Estrin, et. al., "Protocol Independent Multicast Sparse Mode (PIM-SM): Motivation and Architecture", draft-ietf-idmr-pim-arch-04.ps , October, 1996.

[PIM-DM] D. Estrin, et. al., "Protocol Independent Multicast Dense Mode (PIM-DM): Protocol Specification", draft-ietf-idmr-PIM-DM-spec-04.ps, September, 1996.

[PIMMBR] D. Estrin, et. al., "PIM Multicast Border Router (PMBR) specification for connecting PIM-SM domains to a DVMRP Backbone", draft-ietf-idmr-PIMBR-spec-01.ps, September, 1996.

[PIM-SM] D. Estrin, et. al., "Protocol Independent Multicast Sparse Mode (PIM-SM): Protocol Specification", draft-ietf-idmr-PIM-SM-spec-09.ps, October, 1996.

[THALER96] D. Thaler, "Interoperability Rules for Multicast Routing Protocols", draft-thaler-interop-00.ps, November, 1996.

[ESTRIN97] D. Estrin, et. al., "A Dynamic Bootstrap Mechanism for Rendezvous-based Multicast Routing", USC/ISI Technical Report, 1997.

[RFC1636] Braden, R., Clark, D., Crocker, S., and C. Huitema, "Report of IAB Workshop on Security in the Internet Architecture", RFC1636, June 1994.

[RFC1704] N. Haller and R. Atkinson, "On Internet Authentication", RFC1704, October 1994.

[RFC1825] Atkinson, R., "IP Security Architecture", August 1995.

[RFC1826] Atkinson, R., "IP Authentication Header", August 1995.

[RFC1827] Atkinson, R., "IP Encapsulating Security Payload", August 1995.

[RFC1949] A. Ballardie, "Scalable Multicast Key Distribution", RFC1949, June 1996.

[RFC2085] M. Oehler & R. Glenn, "HMAC-MD5 IP Authentication with Replay Prevention", February 1997.

[RFC2104] H. Krawczyk, M. Bellare, & R. Canetti, "HMAC: Keyed Hashing for Message Authentication", RFC2104, February 1997.

[VILLAM95] C. Villamizar, Ravi Chandra, and Ramesh Govindan, "Controlling BGP/IDRP Routing Overhead", draft-ietf-idr-rout-dampen-00.ps, July, 1995.

20. Acknowledgments

Dino Farinacci, Dave Thaler, and Yakov Rekhter provided several insightful comments on earlier versions of this document. Ran Atkinson contributed most of the security ideas contained in this document.

21. Author Information

David Meyer
University of Oregon
1225 Kincaid St.
Eugene, OR 97403
Phone: (541) 346-1747
e-mail: meyer@antc.uoregon.edu