Publication III

Jarno Rajahalme. Incentive-Informed Inter-Domain Multicast. In *INFOCOM 2010, 29th IEEE Conference on Computer Communications Workshops, Global Internet Symposium*, San Diego, CA, USA, March 2010 [167].

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings

Incentive-Informed Inter-Domain Multicast

Jarno Rajahalme
Nokia Siemens Networks

***Abstract*—IP multicast enables reduced link loads for multiple recipients with overlapping data paths. In inter-domain settings, however, the costs and benefits of multicast redundancy elimination fall on different economic entities. Hence, in the absence of multicast specific inter-domain contracts and compensation, IP multicast has remained mostly an intra-domain endeavor.**

We propose *incentive-informed inter-domain multicast*, with which domains base multicast operation solely on their local unicast-derived incentives. Multicast forking will be provided only by domains that find it locally beneficial. Multicast traffic will pass through other domains using unicast paths.

We find that, overall, up to 95% of the optimal IP multicast link load reduction can be attained, varying by the multicast group size. Link loads near the multicast sources can still remain high, mostly due to the considerable uphill diversity in the Internet topology. However, causal dynamics seem to exist for top tier transit providers to offer multicast as a service, significantly reducing the effective uphill diversity, thus enabling highly reduced link loads also on uphill paths. In this case the overall efficiency is also better than with optimal IP multicast.

I. Introduction

Careful examination of current inter-domain traffic incentives reveals that the costs and benefits of IP multicast [1], when applied to inter-domain scenarios, generally fall on different economic entities. This observation enables two possible tracks for inter-domain multicast deployment. Following the lead of [2], we assume that augmenting all existing bilateral traffic contracts with multicast specific accounting [3,4] and processing rules, as required by the current slate of inter-domain multicast protocols [5,6],[1] is not practically possible. Consequently, our focus is on augmenting the multicast service model and protocol operation to be compatible with the existing unicast traffic incentives.

The IP multicast service model was originally defined "without a commercial service explicitly in mind" [8]. Inter-domain networking, however, takes place between largely independent administrative domains, typically representing different economic entities, often in competition with each other. Furthermore, this commercial inter-domain structure has emerged around the unicast service model, where each packet has exactly one sender and one receiver, both financially responsible for their share of the end-to-end connectivity [9]–[11].

In contrast, the essence of the IP multicast model is that it allows sources to send each packet only once, while *the network* delivers the packet to multiple destinations. Assuming the existing unicast contracts between sources and network providers reveals the basic multicast *tussle* (Section II) [12]: The multicast sources are the main benefactors of the service, while their network providers may find their customer traffic levels reduced, decreasing customer revenues, while bearing high next-hop transit costs due to traffic forking to multiple paid transit links.

[1]See [7] for a thorough historical perspective.

The cost/revenue tussle repeats between *all* adjacent domains in the potential multicast delivery paths. Therefore, to enable *any* inter-domain multicast to take place, one domain's decision to *not* participate should not preclude other domains from providing multicast. In the extreme, when no domains provide multicast service, the service should fall back to end-to-end unicast between the sources and destinations.

Expanding on these considerations we propose an *incentive-informed inter-domain multicast* model, where each domain bases its multicast operation on its local unicast-derived incentives. In this model, domains create multicast branching points only for those multicast groups, for which they find forking locally beneficial. All other multicast traffic is forwarded using the normal unicast forwarding tables. The latter is also what happens in domains with no multicast support. Unicast forwarding across domain boundaries requires either changes to the multicast addressing model (as in [13]), or use of tunneling.[2]

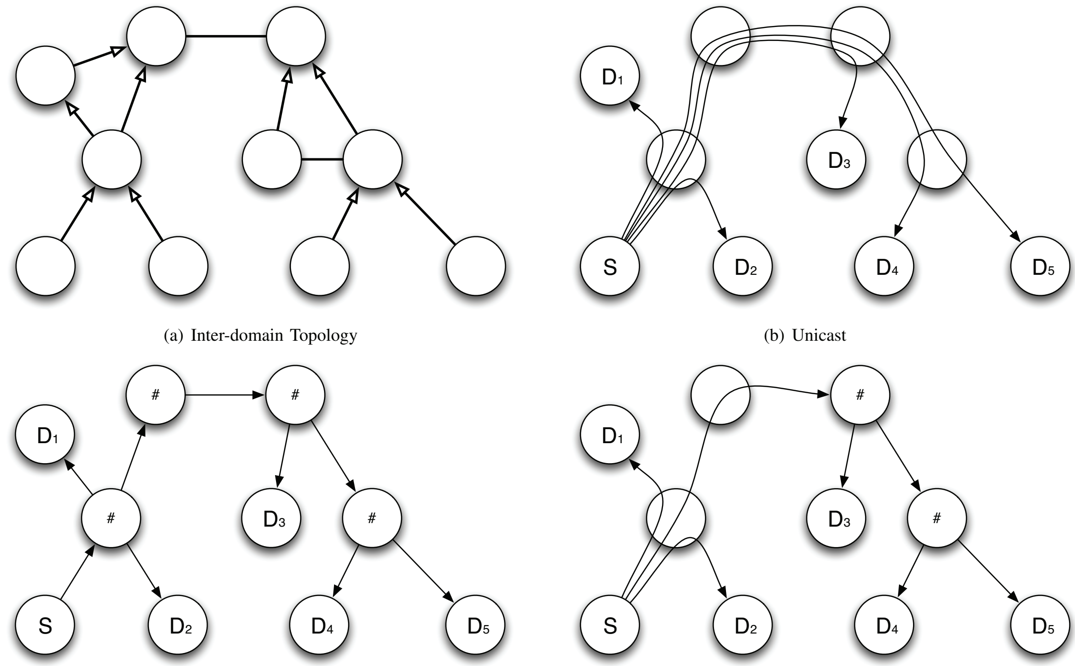
Evaluating the proposed multicast model on the Internet topology (Section V), we find that most of the overall redundancy elimination made possible by IP multicast can be attained by observing local incentives only. Furthermore, if this design is adopted, it appears that top tier transit providers could increase their customer traffic by offering multicast as a service. Evaluating this option (Section VI), we find that the uphill redundancy is almost completely eliminated, and that the operation is more efficient than optimal IP multicast.

II. The Multicast Tussle

The optimal IP multicast delivery structures are typically composed of individual policy-compliant unicast paths, eliminating redundant transmissions on each link (see Figures 1(b) and 1(c)) [3,16]. Noting that the typical unicast traffic policies are derived from the economical incentives of the participating domains, it is perhaps surprising to see that the straightforward combination of unicast paths into multicast delivery trees can be *incentive incompatible*, when only unicast based revenue is considered. For example, in Figure 1(c), for each packet sent by the source S, the next hop provider is sending one packet to another customer, and two to its own providers. Assuming

[2]Unicast tunneling is already part of the current multicast protocols for sending packets from sources to Rendezvous Points (RPs), or from an RP to another [6,14], and as a transitional tool [15].

978-1-4244-6739-6/10/$26.00 ©2010 IEEE



(a) Inter-domain Topology

(b) Unicast

(c) Optimal IP Multicast

(d) Incentive-informed Inter-domain Multicast

Fig. 1. Incentive-informed inter-domain multicast example from source S to destinations D_{1-5}. Customer domains are depicted below their providers. Figure (a) shows also the peering relationships (plain lines). In Figures (b-d) each arrow represents inter-domain traffic, multiple parallel arrows representing redundant transmissions. "#" indicates multicast state management and traffic forking within a domain.

typical volume (tiered) charging, the provider is typically better off *not providing* the multicast service (Figure 1(d)).

It follows, that like peering for unicast traffic [17], inter-domain multicast represents a *tussle* [12] between the stakeholders comprising the Internet. The fact that one party's cost savings may seem like lost revenue to another forms the basic dynamic also in the multicast tussle. Consequently, each preceding transit domain on any given path has an incentive to *block* the following domains from providing multicast service, since enabling multicast down the path leads to reduced upstream customer revenues. Given that the current inter-domain multicast protocols (e.g., [5,6]) depend on the participation of *all* the domains comprising the multicast delivery structures, there is *no design-time resolution* for the multicast tussle.

Thus, the general direction of a deployable multicast solution emerges:

1) downstream multicast operation must not depend on the (protocol) support of any upstream domain,
2) like with IP multicast, the multicast reception request needs to come from the destination, enabling *run-time* tussle resolution, and
3) in each domain, the local *multicast policy*, guiding the protocol operation, must be derived only from the local incentives.

III. Multicast Incentives

In the abstract our aim is simple enough: allow all domains to decide on a case-by-case basis whether they are interested in serving a multicast reception request, or whether such a request should be just forwarded to the next domain towards the source, without taking any action. In order to assess the possible benefits of such a protocol, however, we need to consider the concrete cases where definite decisions can be seen to be made.

A. Assumptions

In the following, we assume that the network service providers are in the business of providing network access and transit services demanded by their customers. Furthermore, we assume the existing unicast contracts, where end customers pay bandwidth tiered flat-rate charges for their Internet access, and transit providers charge per volume, typically on 95 percentile peak basis [17].

The almost universal preference for least cost inter-domain routing, i.e., preferring customer links over peering links, and peering links over provider links, has lead to the *valley-free paths model* [11], in which packets are first forwarded *uphill*, using only provider links. Forwarding is switched to *downhill* portion as soon as possible, maybe over a peering link, after which only customer links are used. More than 99% of Internet paths conform to this model [11]. We use these notions in our analysis below.

For the purposes of comparing multicast and unicast efficacy we assume the destinations to create the traffic demand based on an application level service requested by the end users. The traffic volume received by the ultimate destinations is assumed to be the same in both cases.

B. Local Incentives

Given the above assumptions, we now examine the incentives of all the parties along potential multicast delivery paths, starting from the source:

a) Sources: For sources the benefits of multicast are self-evident: redundancy elimination from the server ports, local links and networks, and most importantly, from the uphill transit links allows both cost savings and the ability to serve larger number of destinations.

b) Uphill Access and Transit Providers: We touched upon this aspect in the Section II above. It follows that assuming only unicast-derived revenue, *all* uphill transit providers are better off *not providing* the multicast service, as their customer revenue is maximized by the multiple unicast transmissions from the source.

c) Downhill Transit and Access Providers: If the transition to downhill portion takes place over a peering link, then the local multicast incentive at the receiving peer depends on the specifics of the peering agreement and the state of the peering link(s). If additional incoming traffic increases costs, or saturates the link(s), providing multicast forking is beneficial.

The transit and access providers closest to receivers, however, are clearly potential beneficiaries due to reduced transit costs. It follows that *when considered separately*, all domains on the *downhill* portions of valley-free paths would benefit from reducing redundancy on their paid transit links. But since their cost savings represent lost revenue from their provider's viewpoint, the only chance for multicast deployment comes through domains unilaterally offering multicast service in cases where they can realize transit cost savings or more customer revenue, or both.

d) Destinations: Destination networks with multiple receivers have a natural incentive for reducing incoming transit traffic. However, following from our traffic demand assumption, the ultimate destinations (hosts, users) of the multicast traffic have little to gain from multicast as such, but there are no obvious disincentives either. The increasing trend of service-specific code running in destination hosts (e.g., client-side scripting) makes the destinations share some of the incentives of the application service providers. This makes it possible for the destination nodes to request multicast service without explicit consent of, nor any harm to, the respective end users.

IV. Domain-level Multicast Policy

Our next step is to derive a generic unicast friendly inter-domain multicast policy – a default set of multicast routing rules that could be safely turned on by all network service providers, and thus make multicast available for all network applications.

Due to the possible negative incentives and additional overhead cause by multicast, the policy needs to include an opt-out mechanism. This consideration calls for unicast forwarding of multicast traffic across domain boundaries.[3]

Since clear multicast incentives exist only at the sources and on the destination side of the inter-domain paths, multicast operation must be requested by each destination, sending a *multicast reception request* towards the source.

We define the policy by specifying domain-level processing rules as follows:

1) **Drop** all requests in violation of the local unicast routing policy. I.e., drop requests coming from provider or peering links, if the next inter-domain hop towards the source would be a non-customer link. This rule enforces valley-freeness of the resulting multicast paths.
2) **Intercept** requests, whose next-hop towards the multicast source would be *either a provider link or a peering or a sibling link for which incoming traffic reduction is desirable*,[4] as follows:
   a) If a local domain-level *multicast branching point* (RP) for this multicast group already exists, the new destination is added and the request is terminated.
   b) Otherwise, local policy is used to determine whether a new RP is created:
      i) Typically an RP would not be created for the first request for a given multicast group, allowing the packets to pass through the domain in unicast mode without any additional overhead. The original request is forwarded towards the source.
      ii) If an RP is created, the request is terminated, and a new request, indicating the new RP as the (only) destination, is sent towards the source.
3) **Forward** all other requests, allowing the requested traffic be delivered without multicast processing in this domain.

Any domain not implementing multicast would follow its normal unicast policy for drop/forward. We omit further state management details for brevity.

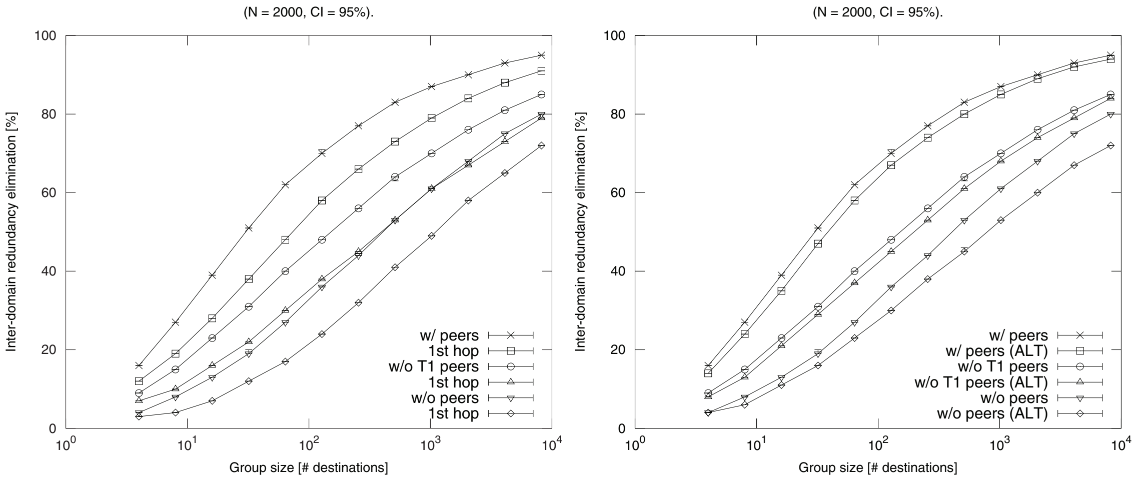
V. Policy Evaluation

The multicast policy given above is sub-optimal by definition, but exactly how much of the optimal multicast efficiency is lost? To answer this question we apply the policy to the observed Internet inter-domain topology [18]. We test the policy operation for different group sizes, as we expect results to differ for small and large groups.

Source and destination domains are taken from power-law distributions derived from the inter-domain transit hierarchy,

[3]Sometimes unicast is preferred also *within* networks with multiple destinations. E.g., in wireless networks, link layer multicast typically requires cell-wide broadcast, using more energy and causing more inter-cell and inter-channel interference, than, say, two energy-optimized unicast transmissions.

[4]Paid peering links would fall into this category. Sibling link [11] processing depends on the type of the next non-sibling relation towards the source.



(a) Efficacy with different multicast peering policies.

(b) Efficacy with and without additional peering links.

Fig. 2. The degree of redundancy elimination with incentive-informed inter-domain multicast. 0% and 100% levels correspond to unicast and IP multicast, correspondingly. "1st hop" refers to the first inter-domain hop from the source for the preceding policy case. (ALT) marks the use of additional peering links.

observing measured utility rank correlations and measured power-law distributions [19]. These variates are re-generated for each simulation round.

The overall level of multicast efficacy can be approximated with the network-wide redundancy elimination metric. However, since the strain of sending multiple unicast flows is highest near the sources, we also look at the number of redundant transmissions on the first hop from the sources.

To overcome some of the contingency in the local policy (Section IV), we observe the efficacy for three different local policies regarding peering links:

1) In the best case, all multicast reception requests, for which the next inter-domain hop towards the source is a peering link, are intercepted, and a local RP is created whenever there are two or more recipients for the multicast group.
2) No multicast peering over tier-1 peering links. Since excessive sending by peers results in settlement payments [9], incoming traffic redundancy is welcomed, and hence is not reduced. This also allows the (competing) peer to keep more of its customer traffic.
3) No multicast redundancy elimination over any peering links. This represents the absolute worst case, as, for example, an access provider peering with a content provider has really no reason to not reduce the overhead in their peering link.[5]

Figure 2(a) shows the mean results[6] of 10 rounds of 200 senders sending multicast traffic to destination groups of different sizes. Graphs for the three policy options regarding traffic arriving over a peering link are shown. For each local policy option both the overall and the first hop redundancy reduction are shown.

The lower level of redundancy elimination on the first hop shows the effect of the lopsided multicast incentives along the end-to-end path. The treatment of incoming peering links is significant, as prior to the peering link there are only customer-provider links, for which we found no unicast compatible incentive for multicast provision.

To assess the effect of the up to 90% of peering links reportedly missing from the CAIDA datasets [20], we augmented the CAIDA topology with 900% additional peering links according to the following simple rules:

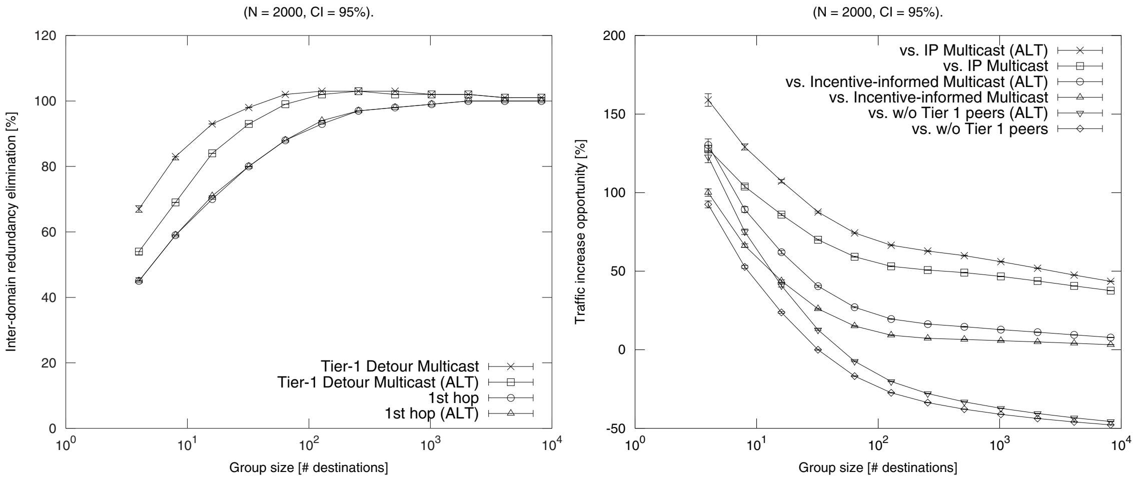
1) All peering relationship at and above the domains with Route Views route monitors [21] are known, so none is added to these domains.
2) No peering for singly-homed stub domains. We assume that domains with interest for peering would first find interest for multi-homing.
3) No peering with (transitive) customers. I.e., no peering with ones providers, their providers etc.

Observing these rules we pick 10 different augmented datasets, each with peering domains selected with our traffic model, so that the domains most likely to originate traffic are also the ones that are more likely to peer than others. As the peering relations are the most dynamic part of the Internet topology, the augmented datasets also provide a view on how the results might change when the topology evolves.

The mean effect of the additional peering is shown in Figure 2(b). For each policy choice defined above, we show the overall efficacy with the CAIDA dataset (same curves as in

[5] In fact, the content provider might *require* the access provider to provide multicast to manage their network load.

[6] The variance in the means is so low that the error bars are barely noticeable.



(a) Tier-1 detour multicast efficacy.

(b) Tier-1 traffic increase opportunity, compared to other forms of multicast.

Fig. 3. The degree of redundancy elimination with tier-1 detour multicast, with incentive-informed inter-domain multicast on the downhill paths. 0% and 100% levels correspond to unicast and IP multicast, correspondingly. “1st hop” refers to the first inter-domain hop from the source for the preceding policy case. (ALT) marks the use of additional peering links.

Figure 2(a)), and the corresponding policy with the augmented datasets. The effect on first hop links is similar (not shown).

While the results above can be considered significantly better than for our multiple unicast baseline, sending hundreds of copies for larger groups still feels wasteful. However, this seems to be the best we can attain with the assumed deployment model. Next we will look for an alternative design with one more assumption, and with strikingly good results.

VI. Tier-1 Detour Multicast

Overall, it seems that *as a block* transit providers would be better off if there was no multicast at all, as making senders use unicast maximizes the overall network traffic and revenue, based on the existing contracts. Luckily Internet transit provision is highly competitive undertaking, making it difficult for any provider to force others to use their services. The steady increase of bilateral peering provides an illuminating example of the dynamic in Internet transit provision. Apart from potential benefits to service quality (e.g., latency), the main motivation for such peering is reduction of paid transit traffic and the corresponding savings from the transit costs. Transit providers face decreasing revenue, but due to competition between transit providers, they can do little about it.

This being the case, it is possible that some transit providers might be able to *draw more customer traffic to their networks by offering multicast as a service*. In such a service sources would include multicast group addresses referring to their service providers in multicast session descriptions, causing multicast reception requests to be routed via their multicast service providers.[7] Sources then use unicast to send the multicast traffic to their multicast service providers, who then fork the traffic to their customer links. This effectively eliminates the use of peering links below the multicast service provider domains in the transit hierarchy. Facing increasing levels of transit traffic, the only sensible move the downhill-side providers can make is to respond to the multicast reception requests they receive from their customers, as described in the preceding sections.

However, such unilateral operation may face a problem with coverage, since increasing the traffic volumes on outgoing peering links may be undesirable. This is because the increased traffic may tilt the peering links out of balance, leading to settlement payments to the peers [9]. To bypass this issue, we assume the sources to send a limited number of copies, one for each tier-1 domain required to cover the given set of destinations, based on the multicast reception requests they receive.

Figure 3(a) shows the mean results of the same simulated multicast cases as reported above in Figure 2(a), but now showing the degree of multicast redundancy elimination with the tier-1 detour design.

Overall, the tier-1 detour option performs better than optimal IP multicast, even when sources still send multiple copies. This is due to the fact that the paths via the tier-1 domains closest to the destinations are on the average shorter than the normal unicast paths inflated by the unicast routing policies [22]. In the case of the largest group size tested (8192) the mean number of copies sent over the first inter-domain hop

[7] Alternatively, various source routing mechanisms could be used for the same effect.

has now reduced to $\sim$11.5. The first hop overhead does not significantly increase even for larger groups, as a relatively small set of tier-1 domains can reach most of the hosts in the Internet.

This kind of policy is clearly attractive whenever the multicast traffic represents new customer traffic for the tier-1 domains. This is not always the case, however. To address this question, we assume that none of the traffic is new, and ask would the tier-1 customer traffic increase, if they offer multicast as a service to *all* Internet users, including some of their paying customers, from which some incoming traffic will be thus *reduced*?

Figure 3(b) shows the degree of traffic increase/decrease, should the tier-1 domains offer multicast as a service. For smaller group sizes the answer is clear, but for larger groups the relative tier-1 benefit depends on the degree of traffic reduction made earlier via the local multicast policy decisions. If all domains consider incoming peering traffic as an expense, then tier-1 detour multicast will beneficial for all group sizes. We also include the comparison to IP multicast, and conclude that tier-1 detour multicast is by far more beneficial for tier-1 transit providers than IP multicast.

Finally, we note that the tier-1 domains seem to be in the unique position where they could host any new bandwidth intensive applications (with or without multicast) and find their customers paying for the increased traffic volumes.

VII. Conclusions

We have observed the incentives at play with inter-domain multicast provision, and found that for inter-domain multicast to be deployable, it must be able to pass through non-participating domains. As the costs and benefits of optimal multicast operation generally fall on different economic entities, most end-to-end paths will have domains with no incentives for multicast provision, when only traffic volume derived revenue is assumed. As a general remark we note that the multicast incentives are unbalanced along the end-to-end paths.[8] Sources and destination-side providers can benefit, while source-side providers may face reduced customer revenues. Destination hosts have no particular incentive for multicast reception per se, if the unicast alternative is available.

Based on these findings we have derived a default *incentive-informed inter-domain multicast policy*. In our evaluation we find that most of the optimal IP multicast redundancy elimination benefits are attainable even when operation is based on local incentives only.

Furthermore, we find that if such a multicast model is adopted, an opportunity to draw some of the lost traffic back to the top tier transit networks opens up. Our evaluation shows that such a *tier-1 detour multicast*, combined with voluntary downhill redundancy elimination according to our default policy, has even higher network-wide efficiency than the optimal IP multicast model. Due to the small number of top tier transit providers being able to reach most of the Internet destinations, the uphill redundancy is also very limited, and scales well also for very large group sizes.

With these findings we conclude that the elusive inter-domain multicast deployment seems possible, given that the inter-domain traffic policies are considered already in the protocol design phase, rather than as an after-thought only.

[8] This mirrors our earlier findings for inter-domain caching incentives [23].

References

[1] S. E. Deering and D. R. Cheriton, "Multicast routing in datagram internetworks and extended LANs," *ACM Trans. Comput. Syst.*, vol. 8, no. 2, pp. 85–110, 1990.

[2] S. Ratnasamy, S. Shenker, and S. McCanne, "Towards an Evolvable Internet Architecture," in *ACM SIGCOMM'05. Proceedings*, 2005, pp. 313–324.

[3] J. Feigenbaum, C. Papadimitriou, and S. Shenker, "Sharing the Cost of Multicast Transmissions," *Journal of Computer and System Sciences*, vol. 63, no. 1, pp. 21–41, 2001.

[4] J. C.-I. Chuang and M. A. Sirbu, "Pricing Multicast Communications: A Cost Based Approach," in *Proceedings of INET*, 1998.

[5] T. Bates, Y. Rekhter, R. Chandra, and D. Katz, "Multiprotocol Extensions for BGP-4," IETF, RFC 2858, Jun. 2000.

[6] B. Fenner and D. Meyer, "Multicast Source Discovery Protocol (MSDP)," IETF, RFC 3618, Oct. 2003.

[7] K. C. Almeroth, "The Evolution of Multicast: From the MBone to Inter-Domain Multicast to Internet2 Deployment," *IEEE Network*, vol. 14, no. 1, pp. 10–20, Jan. 2000.

[8] C. Diot, B. Levine, B. Lyles, H. Kassem, and D. Balensiefen, "Deployment issues for the IP multicast service and architecture," *Network, IEEE*, vol. 14, no. 1, pp. 78–88, Jan/Feb 2000.

[9] W. B. Norton, "Internet Service Providers and Peering," Unpublished draft, Tech. Rep., 2001.

[10] G. Huston, "Interconnection, Peering, and Settlements," in *Proc. Internet Global Summit (INET)*, Jun. 1999.

[11] L. Gao, "On inferring autonomous system relationships in the Internet," *Networking, IEEE/ACM Transactions on*, vol. 9, no. 6, pp. 733–745, Dec 2001.

[12] D. Clark, J. Wroclawski, K. Sollins, and R. Braden, "Tussle in Cyberspace: Defining Tomorrow's Internet," *Networking, IEEE/ACM Transactions on*, vol. 13, no. 3, pp. 462–475, June 2005.

[13] I. Stoica, T. Ng, and H. Zhang, "REUNITE: A Recursive Unicast Approach to Multicast," in *IEEE INFOCOM*, vol. 3, 2000, pp. 1644–1653.

[14] B. Fenner, M. Handley, H. Holbrook, and I. Kouvelas, "Protocol Independent Multicast – Sparse Mode (PIM-SM): Protocol Specification (Revised)," IETF, RFC 4601, Aug. 2006.

[15] H. Eriksson, "MBONE: The Multicast Backbone," *Commun. ACM*, vol. 37, no. 8, pp. 54–60, 1994.

[16] C. Chow, "On Multicast Path Finding Algorithms," in *INFOCOM'91. Proceedings*. IEEE, 1991, pp. 1274–1283.

[17] P. Faratin, D. Clark, P. Gilmore, S. Bauer, A. Berger, and W. Lehr, "Complexity of Internet Interconnections: Technology, Incentives and Implications for Policy," in *TPRC. Proceedings*, 2007.

[18] "The CAIDA AS Relationships Dataset," http://www.caida.org/data/active/as-relationships/, Aug. 10th 2009.

[19] H. Chang, S. Jamin, Z. Morley, and M. W. Willinger, "An Empirical Approach to Modeling Inter-AS Traffic Matrices," in *ACM SIGCOMM IMC'05. Proceedings*, 2005, pp. 139–152.

[20] R. V. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "In Search of the Elusive Ground Truth: The Internet's AS-level Connectivity Structure," *SIGMETRICS Perf. Eval. Rev.*, vol. 36, no. 1, pp. 217–228, 2008.

[21] "Route views peers," http://www.routeviews.org/peers/, Jul. 2009.

[22] S. Savage, T. Anderson, A. Aggarwal, D. Becker, N. Cardwell, A. Collins, E. Hoffman, J. Snell, A. Vahdat, G. Voelker, and J. Zahorjan, "Detour: Informed Internet Routing and Transport," *Micro, IEEE*, vol. 19, no. 1, pp. 50–59, Jan/Feb 1999.

[23] J. Rajahalme, M. Särelä, P. Nikander, and S. Tarkoma, "Incentive-Compatible Caching and Peering in Data-Oriented Networks," in *ReArch'08. Proceedings*, 2008.