On the incremental deployment of overlay information centric networks

Konstantinos KATSAROS, Charilaos STAIS, George XYLOMENOS, George C. POLYZOS
Athens University of Economics and Business, Patision 76, Athens, 104 34, Greece
Tel: + 30 210 8203693, Fax: + 30 210 8203686
Email: ntinos@aueb.gr, stais@aueb.gr, xgeorge@aueb.gr, polyzos@aueb.gr

***Abstract*— It has long been realized that the proliferation of information-centric applications and services must be reflected in a corresponding shift of the underlying Internet architecture. Even though users increasingly focus on the desired information, the underlying network still focuses on the endpoints providing/consuming this information and in many cases this mismatch has resulted in an inefficient utilization of network resources, as demonstrated by peer-to-peer (P2P) and file sharing applications. In view of this situation, many research projects have focused on the investigation of alternative networking models centered around information. However, less attention has been paid to the transition process from the current end host centric model to an information centric one. In this paper, we propose an overlay multicast-enabled, publish-subscribe architecture and focus on its gradual deployment both inside administrative domain boundaries as well as across the Internet. Our simulation results demonstrate the benefits for individual network operators as they gradually adopt our new networking model, and shed further light on the extent of deployment required within an administrative domain in order for our approach to perform optimally.**

***Index Terms*— Content centric, multicast, publish-subscribe, content distribution.**

I. INTRODUCTION

The Internet protocols were originally designed to exchange traffic between pairs of communicating end hosts, following the prevailing communication patterns of previous networks. Communication patterns have evolved since then, and Internet use has shifted towards *information-centric* services and applications, such as content delivery networks (CDNs), cloud computing services and peer-to-peer (P2P) file sharing. In these services, in sharp contrast to the underlying Internet model, the focus is on the information itself, rather than on the end hosts producing or consuming it. Hence, these services are implemented as overlays on top of the information-agnostic network substrate [1].

While the development of overlays has in many cases yielded important benefits to end users, it has also reviled the problems incurred by the mismatch between the prevailing communication patterns and the underlying network model. This is especially evident in the case of P2P file-sharing applications, in which end users benefit in terms of perceived download times and content providers decrease their resource requirements. However, the use of the (inappropriate) end-to-end communication model makes these applications plague the Internet with redundant unicast transmissions, e.g. many nearby nodes independently downloading the same data from a faraway node [2].

In view of this situation, in our previous work [3] we proposed the deployment of a *router assisted overlay multicast* (RAOM) architecture for information dissemination, based on the Scribe overlay multicast [4] and the Pastry overlay routing [5] schemes. The publish/subscribe nature of RAOM decouples information producers from consumers, thus breaking the current end-to-end model. The deployment of overlay functionality inside the network instead of at the edges, as proposed by RAOM, leads to substantially improved multicast tree properties and reduced signaling overhead [3].

In this paper we focus on the transition to an ubiquitous RAOM deployment, by investigating the properties of the resulting overlay multicast trees in two directions. First we examine the benefits that would motivate individual network operators to invest on the RAOM provision, by comparing their perceived performance against operators that do not engage in RAOM. Second, we investigate the extent of the investment required inside domain boundaries in order to provide optimal performance, by studying the resulting tree properties for various deployment density degrees.

The remainder of this paper is organized as follows. In Section 2 we present the RAOM architecture, while in Section 3 we discuss its incremental deployment. Section 4 first provides a thorough description of the simulation environment and then presents and discusses the results obtained. We conclude in Section 5.

II. THE RAOM ARCHITECTURE

The unsuitability of the Internet for information centric applications is evident in its lack of a multicast facility: multicast is inherently information-centric, it decouples the producers and consumers of information, and it promotes the efficient use of network resources. However, while IP multicast has been available for more than a decade, it has not been widely adopted for various reasons. IP multicast routing does not scale well: unlike unicast addresses that can be easily aggregated, nearly identical multicast addresses refer to completely different member sets, therefore routers must allocate memory and perform signaling separately for each group. There are also no gains to be made by supporting multicast, unless if all routers support it, therefore there are no incentives for individual routers to start doing so. Finally, the lack of group management facilities for access control and billing, have also contributed to the limited deployment of IP multicast [6]. This is largely due to the fact that group management is inherently correlated with application logic, which cannot be incorporated into the network layer IP multicast model. Similarly, higher layer functions such as error recovery, congestion and flow control have not been properly dealt with in IP multicast.

Our RAOM architecture takes these factors into account, first, by employing a highly scalable application layer multicast scheme, and second, by pushing required functionality inside access networks. As a result, scalable multicast is achieved without relying on the ubiquitous deployment of overlay functionality. In addition, by operating at the application layer RAOM enables the exploitation of already available solutions regarding error, flow and congestion control. Finally, the proxy character of the established overlay nodes (described in Section II-B) facilitates group management, as these nodes may act as intelligent gateways of end hosts to the multicast substrate.

A. Overlay multicast with Scribe

The absence of IP multicast support has led to the emergence of overlay solutions that do not require network support, such as those based on *Distributed Hash Table* (DHT) substrates. In DHT substrates like Pastry [5] a uniform identifier space is distributed among the participating nodes; these are regular end hosts that use the underlying IP transport transparently to the routers. The advantage of such schemes is that the amount of routing state required per node and the maximum number of hops required to reach any other node scale logarithmically with the number of nodes, a critical feature for Internet scale systems. On the other hand, packets following overlay routes do not take the shortest path towards their destination. By employing proximity metrics, such as the number of IP hops or the round trip time, Pastry attempts to minimize this side-effect by taking network locality into account: among the possibly many DHT nodes that are closer to a packet's identifier, and which could thus continue relaying the data, Pastry chooses the closest one with respect to the employed proximity metric.

Scribe [4] supports multicast distribution over a DHT substrate by mapping the name of each group to an identifier and making the node responsible for that identifier the group's *rendezvous* (RV) point. Receivers join the group by sending a join message towards the group identifier; as the message propagates towards the RV point, reverse path routing state is established until a node already in the tree is found, thus forming a multicast tree rooted at the RV point. A sender simply routes data towards the group identifier, so that the RV point may then propagate it over the established tree.

The reliance of Scribe on end hosts may however lead to inefficiencies. An end host that is an interior node in some trees will limit the bandwidth available to all those trees to that of its access link. This can be avoided by exploiting the properties of the underlying DHT to create a set of trees such that each node will be an interior node for only one of them [7]. However, this solution is tied to a specific overlay routing scheme (in this case, Pastry). In addition, an end host that is an interior node even for a *single* tree, may still be a bottleneck: as shown in Figure 1(a), data in transit has to enter and exit the RV point (peer *b*) and the other two internal end nodes *c* and *d*, only one of which (peer *c*) is also a receiver, via their access links. If these access links are asymmetric, the tree bandwidth will be limited by the, typically lower, uplink bandwidth. Another issue is that neighboring end hosts may download the same content via separate tree branches, thus incurring unnecessary network load. For example, in Figure 1(a) the two peers *e* and *f* receive separately the content from their parent in the tree (peer *d*). Note that in this example peers *b* and *d* act as intermediate tree nodes without being receivers. This arises when nodes participating in various multicast groups share the same DHT substrate, so as to share maintenance costs between groups and improve routing performance by increasing the available overlay paths.

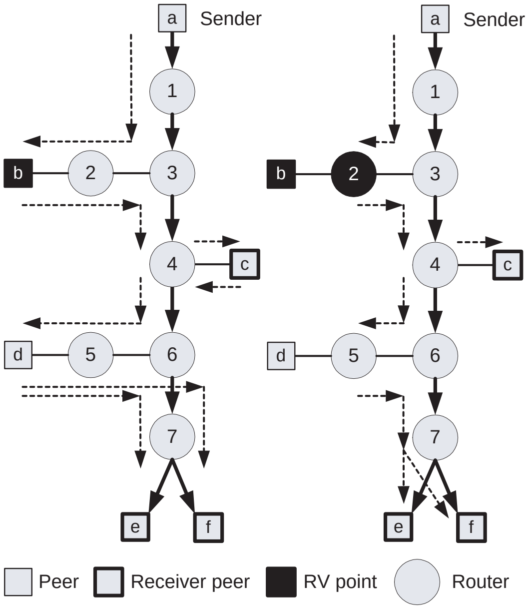


Fig. 1. Overlay multicast: (a) non router assisted (b) router assisted.

B. Router assisted overlay multicast

In order to create an information-centric network fabric that avoids the above problems, in [3] we proposed using the access router of an end host as its *proxy* in the DHT substrate *and* in the overlay multicast scheme. This means that the access router participates in the DHT on behalf of the attached peer. If multiple peers are attached to the same access router, a single place will be held by it in the DHT, that is, the access router will always be assigned a single portion of the identifier space. Similarly, the access router acts as a proxy for end hosts in the Scribe trees, that is, the router is responsible for joining the multicast groups indicated by the attached end hosts and forwarding traffic to them. The access router may also participate in a multicast tree as an interior node, subject to its position in the identifier space and the operation of regular Scribe, that is, if in regular Scribe its attached hosts were interior nodes of that tree. In this case, it also forwards the incoming traffic to its tree descendants.

The proposed proxy role of access routers presents some significant advantages regarding the characteristics of the created distribution trees. First, as shown in Figure 1(b), data do not need to cross the access links of interior tree nodes at all, only crossing the, typically faster, downlink direction of those access links leading to nodes that are members of the group. For example, data will not cross the access link between peer *d* and router *5* at all, and it will only cross the access link between peer *c* and router *4* in the downlink direction so as to deliver the content to peer *c*. Second, multiple tree branches towards end hosts attached to the same access router can be aggregated in a single branch leading to that access router (in our example router *7*). For the entire distribution tree of Figure 1, router assistance means that a packet transmitted to the group will only cross 12 instead of 20 links with regular Scribe (or 8 with an optimal IP multicast tree), avoiding the uplink direction of access links. Therefore, in router assisted overlay multicast the paths through the distribution trees become shorter and faster, while redundant transmissions over the access links of intermediate nodes are avoided, preventing the (normally slower) uplinks from becoming bottlenecks.

C. Mobility support

With the proliferation of mobile devices, any new network architecture must inherently support mobility. Our architecture builds on the inherent capability of multicast to support mobility by localizing routing updates incurred by node movement. In our *Overlay Multicast Assisted Mobility* (OMAM) scheme [8], [9], each mobile node is served by one of the RAOM-enabled access routers (henceforth called *overlay access routers* (OARs)) deployed in the mobile access network, sending and receiving data via appropriately established multicast trees. In dense OAR deployments, a change of location may result in the mobile node attaching to the same multicast tree through another OAR residing at its new location. Based on the routing convergence property of Pastry, it is expected that the new subscription will reach a common ancestor of the previous and the current serving OARs in a network distance approximately equal to the distance traveled during handoff, which is usually short. In [8], [9], it was shown that localizing routing updates has the potential to provide lower handoff delays than Mobile IPv6.

III. INCREMENTAL DEPLOYMENT

The benefits stemming from router assistance were investigated in [3], by comparing the performance of the overlay when all access routers operate as proxies of their attached end hosts, against the case where each end host participates in the overlay by itself. It was shown that RAOM achieves a significant reduction of both path stretch (up to 45%) and link stress (up to 19%), as well as an improvement in overlay network establishment and maintenance overhead (up to 45%), at the cost of increased forwarding overhead for the access routers. The question we pose in this paper is how RAOM would perform if only a fraction of the access routers participated in the overlay as proxies for their end hosts. We have decided to investigate these issues by studying the incremental deployment of RAOM in Interdomain and Intradomain scenarios, as described below.

The *Interdomain* scenario studies the adoption of the proposed architecture by individual network operators. Since a synchronized Internet wide deployment is infeasible, it is important to investigate the performance of RAOM as it gradually gets deployed across the Internet. Specifically, we examine whether RAOM is beneficial even if it is not universally adopted by all network operators, as well as whether there exist performance related incentives for a single network operator to first adopt RAOM.

The *Intradomain* scenario studies the deployment density of the proposed architecture inside the boundaries of each administrative domain. Since RAOM requires the deployment of additional network infrastructure to provide overlay functionality on behalf of the hosts, it is important to examine the relationship between the size of that investment and the perceived gains for the network operator.

IV. EVALUATION

A. Simulation platform

In order to thoroughly study the progressive deployment of the proposed architecture, we have performed an extensive set of simulations based on the implementations of Pastry and Scribe provided by OverSim [10], an overlay network simulation framework for the OMNeT++ simulation environment [11]. In our simulations we employed *transit-stub* topologies produced by the Georgia Tech Internet Topology Model (GT-ITM) [12]. The created topologies consisted of 24 backbone routers in 8 transit domains. Each backbone router supports a single stub domain with 50 access routers on average.

Regarding multicast groups and their sizes, a Zipf-like distribution was used for the size of each group, that is, the r-th group had a size equal to $\lfloor Nr^{-1.25} + 0.5 \rfloor$, where N was the total number of overlay nodes, as in [4]. We focus on dense topologies with 2400 users participating in 75 groups. The members of each group were randomly selected from the entire end host population, meaning that each end host may have participated in many groups. For each group, a random identifier was chosen and a *non-member* end host was randomly selected as the sender. In all scenarios, the target number of end hosts were attached to randomly chosen stub routers. In networks without RAOM support, end hosts entered the overlay themselves, while in networks with RAOM support, each end host was proxied by the closest proxy access router.

B. Simulation Results

Our evaluation focuses on the gains and losses of individual *Autonomous Systems* (AS) in terms of the degree of RAOM adoption. We employed the following metrics, both measured on a per AS basis:

- *Path Length*. It refers to the number of hops traversed by data packets from the root of a tree until the end hosts. It is expressed as the average length of all paths terminating at end hosts within each AS, across all trees. It aims to capture the impact of RAOM on end user experience. Shorter path lengths are expected to increase end-user satisfaction, thus providing an advantage to a specific network operator, against operators with longer delivery paths.

- *Transmission Stress*. It refers to the total number of hop by hop data packet transmissions inside an AS's boundaries for the delivery of a single data packet via all trees, divided by the number of end users in that AS. It aims to capture the total traffic load imposed on an AS normalized by the number of end hosts supported. It includes both transmissions required for packet delivery to end hosts within the AS, as well as transmissions required to serve hosts in other AS's.

In order to evaluate the incremental deployment process described above, we created three basic scenarios that aim to capture the behavior of RAOM across the Interdomain and Intradomain deployment dimensions. These are described in the following sections.

1) Initial deployment incentives: In the first scenario we focus on the initial deployment phase where a single network operator adopts RAOM in its AS. The aim of this scenario is to investigate the potential benefits for this AS compared to other ones, and whether these would provide incentives for the adoption of RAOM. As we can see in Figure IV-B, even with a relatively sparse deployment, the performance perceived at a single RAOM AS is substantially improved. More specifically, we notice an average reduction of 41% and 44% in path length and transmission stress respectively. In effect, we can see that individual operators have strong incentives to adopt RAOM, as both the traffic load on their network is reduced and the delivery of content to their customers is substantially improved.

2) Interdomain incremental deployment: In this scenario we take a step further and investigate the performance perceived by AS's as they progressively engage in RAOM. Each participating AS enhances 50% of its routers with the proposed functionality. Figure IV-B.1 shows that as RAOM is adopted by increasingly more AS's, the performance perceived in these AS's is consistently better than the one perceived at AS's not supporting RAOM. In RAOM supporting AS's, data delivery paths are from 28% up to 40% shorter, while the networks of their non RAOM supporting counterparts' experience 43% to 57% higher traffic. Hence, after the initial adoption by a single AS, the incentives for deploying RAOM are still preserved throughout the progressive deployment process.

3) Intradomain incremental deployment: In this scenario we investigate the impact of deployment density inside each network in the case of ubiquitous RAOM deployment, that is, when all ASs have deployed RAOM. Figure IV-B.2 presents the performance perceived for different deployment densities. As the deployment level increases, so does the transmission stress on the AS's, without visible improvements on path length though. By increasing the number of deployed overlay entities in each AS, we make the overlay routing substrate denser, therefore Scribe trees become denser, with more branches at each level, resulting in higher transmission stress. Therefore, a deployment level of 25% is perfectly sufficient for RAOM.

V. CONCLUSIONS

In this paper we have studied the evolutionary process towards an overlay, information-centric network architecture. Our results demonstrate that network operators have strong incentives to adopt the proposed RAOM architecture, such as significant gains in terms of transmission load and end user satisfaction. Even after initial deployment, RAOM supporting network operators have a consistent advantage over non RAOM supporting ones. Our results also show that sparse intradomain RAOM deployments are sufficient, indicating that an operator can reap the benefits of RAOM by deploying it at only 25% of its routers, regardless of what other operators do.

REFERENCES

[1] D. Clark, B. Lehr, S. Bauer, P. Faratin, R. Sami, and J. Wroclawski, "Overlay networks and the future of the Internet," *Communication & Strategies*, vol. 63, 2006.

[2] T. Karagiannis, P. Rodriguez, and K. Papagiannaki, "Should Internet service providers fear peer-assisted content distribution?," in *Proc. of the Internet Measurement Conference*, pp. 63–76, 2005.

[3] K. Katsaros, N. Bartsotas, and G. Xylomenos, "Router assisted overlay multicast," in *Proc. of the Euro-NGI Conference on Next generation Internet Networks*, pp. 329–350, 2009.

[4] M. Castro, P. Druschel, A.-M. Kermarrec, and A. Rowstron, "SCRIBE: A large-scale and decentralized application-level multicast infrastructure," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 8, pp. 100–110, 2002.

[5] A. Rowstron and P. Druschel, "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems," in *Proc. of the Middleware Conference*, pp. 329–350, 2001.

[6] C. Diot, B. N. 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, 2000.

[7] M. Castro, P. Druschel, A.-M. Kermarrec, A. Nandi, A. Rowstron, and A. Singh, "SplitStream: High-bandwidth multicast in cooperative environments," in *Proc. of the Symposium on Operating Systems Principles*, pp. 298–313, 2003.

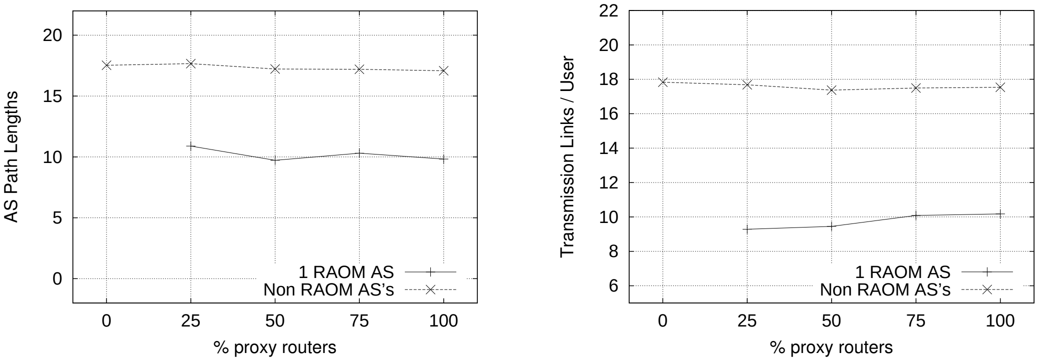
[8] K. Katsaros, N. Fotiou, G. Polyzos, and G. Xylomenos, "Overlay Multicast Assisted Mobility for Future Publish/Subscribe Networks," in *Proc. of the ICT Mobile Summit*, 2009.

[9] K. Katsaros, N. Fotiou, G. Polyzos, and G. Xylomenos, "Supporting Mobile Streaming Services in Future Publish/Subscribe Networks," in *Proc. of the Wireless Telecommunications Symposium (WTS)*, 2009.

[10] I. Baumgart, B. Heep, and S. Krause, "OverSim: A flexible overlay network simulation framework," in *Proc. of the IEEE Global Internet Symposium*, pp. 79–84, 2007.

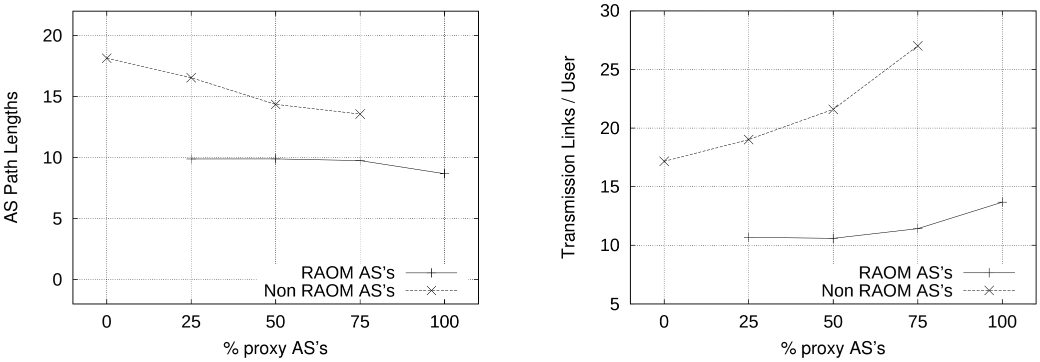
[11] A. Varga, *OMNeT++ Simulator Home Page*. http://www.omnetpp.org.

[12] E. Zegura, K. Calvert, and S. Bhattacharjee, "How to model an internetwork," in *Proc. of the IEEE INFOCOM*, vol. 2, pp. 594–602, 1996.



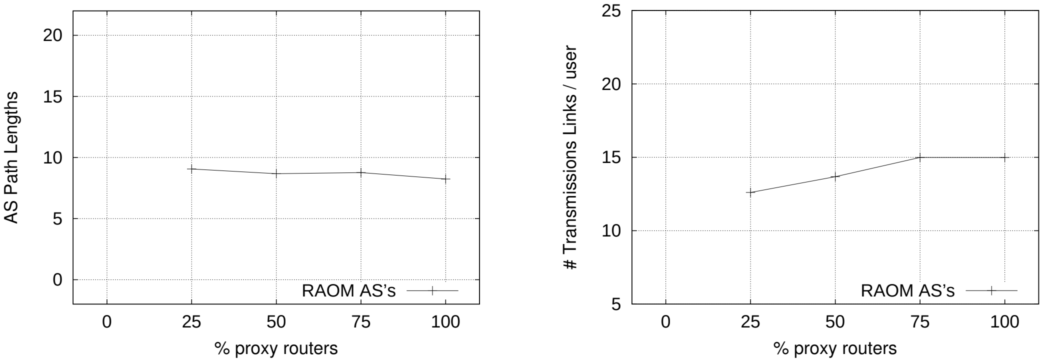
(a) Path Length (b) Transmission stress

Fig. 2. 1 RAOM enabled AS vs. Non RAOM enabled AS's.



(a) Path Length (b) Transmission stress

Fig. 3. Interdomain incremental deployment (50% Intradomain deployment density).



(a) Path Length (b) Transmission stress

Fig. 4. Intradomain incremental deployment (100% Interdomain deployment density).