OASIS: An Overlay Abstraction for Re-architecting Large Scale Internet Group Services

Matthias Wählisch[1,2], Thomas C. Schmidt[2], and Georg Wittenburg[1]

[1] Freie Universität Berlin, Institut für Informatik, Takustr. 9, 14195 Berlin, Germany
[2] HAW Hamburg, Department Informatik, Berliner Tor 7, 20099 Hamburg, Germany
{waehlisch,t.schmidt}@ieee.org, wittenbu@inf.fu-berlin.de

Abstract. There is an increasing economic desire driven by widespread applications like IPTV or conferencing that a next generation Internet will grant transparent group communication service to all its stationary and mobile users. In this paper, we present a generic approach to inter-domain multicast, which is guided by an abstract, DHT-inspired overlay, but may operate on a future Internet architecture. It is based on the assumptions of a globally available end-to-end unicast routing between resolvable locators, taken from a name space that allows for aggregation. Our protocol design accounts for this aggregation, leading to forward-path forwarding along bidirectional shared distribution trees in prefix space. The scheme facilitates multipath multicast transport, offers fault-tolerant routing, arbitrary redundancy for packets and paths and remains mobility agnostic. We present OASIS, its application to IPv6, and evaluate signaling costs analytically based on its k-ary tree structure.

Keywords: Prefix-directed multicast, bidirectional shared tree, Internet architecture, IPv6.

1 Introduction

The idea to extend unicast capabilities by a multicast group service already arose, when the Internet was still in its early, premature state of development [1]. Multicast communication techniques have been under debate since Deering introduced the host group model to the Internet layer [2]. Until today, the initial approach of Any Source Multicast (ASM) routing remained hesitant to spread beyond limited, controlled environments. A fundamental dispute arose on multicast concepts in the end-to-end design concept, questioning the appropriate layer where group communication service should reside on. For several years, the focus of the research community turned towards application layer multicast. Many P2P concepts and solutions of different kind have been developed. We limit reference to the prominent, structured schemes Scribe [3], Bayeux [4] and multicast on CAN [5].

Cost efficiency arguments, however, repeatedly stimulated multicast to be re-settled on the lowest possible layer, exemplified by a commonly available point-to-multipoint service in current wired and wireless transmission technologies.

Meanwhile, new demands for group communication are arriving with increasing intensity, e.g., multimedia streaming (IPTV), large-scale gaming (MMORPGs) and conferencing in mobile environments, or service discovery and configuration in service-oriented architectures. But a fundamental assumption common to multicast routing, the symmetry of connectivity and routing, has changed [6]. In many cases, current multicast trees derived from reverse-path forwarding (RPF) do not lead to efficient or policy-compliant paths.

The Internet has been designed as an overlay abstraction to interconnect networks and enable communication between heterogeneous devices. Guided by an end-to-end design principle [7], transport intelligence is located at edges, while IP provides a common utility of uniform addressing and global unicast routing [8,9].

Limitations in scalability and flexibility, mainly due to shortfalls in addressing, naming and binding [10], but also a current state of deployment that largely deviates from the original IP model [6], led to an ongoing debate on re-architecting the Internet and gave rise to several proposals of network decomposition. Examples include the meta-principle [11], network pluralism in Plutarch [12], and autonomous domains in TurfNet [13]. At the same time, new structures of addressing and routing arose on the application layer [14], resulting in Distributed Hash Tables (DHTs) like Chord [15], Pastry [16] and CAN [17]. Even though paradigmatically unrelated, these two classes of distributed overlay architectures maintain *globally resolvable identifiers* in combination with a *binding to an end-to-end routing*. More importantly, all designs – explicitely or implicitely – rely on *aggregation mechanisms for locators* to obtain scalable routing information bases, thereby addressing a central long-term concern of the IETF. In this work, we start from those three assumptions and construct a group communication mechanism that is based on the aggregation principle.

This paper presents a generic multicast forward-path forwarding approach, which is guided by an abstract, DHT-inspired overlay, but may operate on a future Internet architecture. Such routing may reside on the network or application layer, and may be transparently mapped into a hybrid architecture that procures for prefix continuation. Without loss of generality, we denote locators as part of a prefix space. Within this summable prefix space, the group overlay is defined according to prefix branch points, which will correspond to the topological network structure, provided proximity-aware numbering or routing is in use.

This prefix-directed multicast gives rise to source-specific bidirectional shared distribution trees, which lead to forward-directed packet distribution along shortest paths for any source. Receiver subscriptions, though, follow the general ASM model. On the price of an enhanced signaling load for small groups, this scheme offers fault-tolerant routing, arbitrary redundancy for packets and paths and remains mobility agnostic. Furthermore, it facilitates dynamic multipath transport without additional effort and may give rise to end-to-end resource pooling in multicast, thereby filling the gap left in the recently published research agenda [18].

The remainder of this paper is structured as follows: We introduce prefix directed multicast in section 2. The core protocol can be adapted to the Internet layer. This is shown exemplarily for IPv6 in section 3. Based on its structural

design, all major performance properties can be evaluated analytically, which allows to pre-calculate network costs. Results are presented in parts in section 4, followed by conclusions and an outlook in section 5.

2   Prefix-Directed Multicast

In this section, we present the prefix-directed abstraction for multicast, which enables an arbitrary node to distribute data along forward-oriented source-specific paths. Multicast branch establishment is triggered by ASM receiver joins. Using an algorithm inspired by structured overlays, multicast participants jointly construct a bidirectional shared tree in prefix space. Disregarding specific topology concepts, we explain the basic protocol from the perspective of equal 'peers', which in turn may take the role of designated routers or hybrid gateways in a deployment case as described in section 3. The scheme does not rely on any kind of rendezvous point or bootstrapping and operates directly on top of a unicast routing scheme for locators that allow for aggregating prefixes. It exhibits strictly predictable costs, which scale logarithmically with the number of receivers.
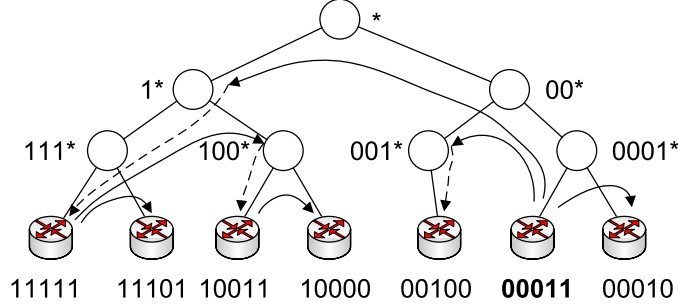
2.1   The Protocol

The key operation of the multicast protocol is to construct a prefix-based distribution tree, in which a leaf is labelled with the locator ID of a multicast listener. Multicast branching is performed at inner vertices, labelled with the longest common prefix of its children. Each inner vertex can be mapped to any peer, whose ID shares the prefix label. The binding of a peer to a prefix will be done according to the local unicast routing information base. Mapping to an efficient topological delivery is thus derived from the deployed routing.

The prefix structure is assumed to reflect the aggregation principle and immediately gives rise to a structured broadcast (see [19]). A broadcast prefix tree can be instantaneously constructed by identifying all peer IDs as leaves. For sending a packet from the root to the leaves of the broadcast prefix tree, each peer needs to decide on packet replication according to its current branching position on the tree. This context awareness can be gained from adding a *destination prefix* $\mathcal{C}$ to the packets, which will be hop-wise updated with growing length. Forwarding is then simply achieved by routing to all neighboring prefixes that share $\mathcal{C}$. This PREFIX FLOODING can be applied at any level of the tree structure (cf. figure 1).

In contrast to broadcast, multicast implements a selective distribution strategy, in which final receivers represent a subset of peers. Any sender, which itself forms a leaf in the prefix tree, will 'shift' the packet up to the (virtual) root of the tree and initiate a forwarding according to prefixes populated by receivers. In this way, the prefix tree is bidirectionally traversed.

All prefix-directed multicast peers will derive semantically identical trees, but will hold only a selected, location-dependent knowledge thereof. Routing correspondences are to be extracted from unicast routing tables and thereby differ from node to node. Multicast forwarders need not memorize the entire group

Fig. 1. Multicast forwarding from source 00011 to prefix neighbors (straight arrows) in a binary prefix tree. Dashed arrows indicate actually selected peers.

specific multicast tree, but will only be required to persist the prefix neighbors of all associated vertices. We call a vertex label to be *associated with* the node if the label represents a prefix of the peer locator ID.

Group Membership Management. Each peer is a potential multicast forwarder, serving as an intermediate destination for a prefix it shares. Consequently, a new multicast receiver has to be announced so that all forwarding nodes can store the corresponding neighboring prefix. This prefix neighbor represents the root of a subtree, which subsumes multiple multicast listeners. Thus, only the first join and last leave needs to be propagated outside this subtree.

To distribute data along a multicast distribution tree, a multicast peer K with locator ID $\mathcal{K}$ maintains an (interface-independent) multicast forwarding table for each multicast group. This list contains all prefixes, which serve as destinations adjacent to K. For a group G, we denote the multicast forwarding table by MFT_G. For prefixes $\mathcal{L}$ and $\mathcal{L}'$, we denote the longest common prefix by $LCP(\mathcal{L}, \mathcal{L}')$ and its length by $|\mathcal{L}|$.

To join or leave a multicast group, a multicast peer injects a state update into the unicast prefix tree. The first and last receiver of the group flood their join and leave message in the complete (unicast) network. For all further group members, the state update is propagated within the smallest subtree including some receiver and covering the new multicast listener. The algorithm works as follows:

JOIN/LEAVE INJECTION

  ▷ Invoked at peer $\mathcal{K}$ for group G
1 **if** $MFT_G = \emptyset$
2   **then** PREFIX FLOODING Join/LeaveMessage To *
3   **else** Select $\mathcal{L} \in MFT_G : |\mathcal{L}| \geq |\mathcal{L}'|, \forall \mathcal{L}' \in MFT_G$
     ▷ Creates root of subtree to flood
4     $\mathcal{C} \leftarrow LCP(\mathcal{L}, \mathcal{K})$
5   PREFIX FLOODING Join/LeaveMessage To $\mathcal{C}$

On the reception of a multicast state update, a peer determines its longest common prefix $\mathcal{L}$ with the originator $\mathcal{P}$, extracts the prefix of $\mathcal{P}$ with length $|\mathcal{L}| + 1$, and adds (or removes) it to its multicast forwarding table. Thereafter the message is transmitted to the downtree neighbors in prefix space.

As prefix lengths are extended in each forwarding step, this algorithm obviously terminates. Furthermore, from an induction on the number of group members follows that the multicast join algorithm constructs a spanning tree at each peer covering all receivers, provided the unicast prefix routing table is complete at all nodes.

Data Dissemination. Based on its group membership functions, prefix-directed multicast constructs a bidirectional shared tree covering all overlay multicast listeners. The prefix neighbors that represent receivers are stored in a decentralized multicast forwarding table MFT_G, which is controlled individually by each forwarding node. Any arbitrary peer can act as multicast source by sending data to all entries in its MFT_G. The packets will then be forwarded to the leaves of the multicast tree. Conceptually this corresponds to the PREFIX FLOODING approach, whereas branching is guided by the multicast forwarding table:

PREFIX MCAST FORWARDING

$\quad\quad \triangleright$ On arrival of packet with destination prefix $\mathcal{C}$
$\quad\quad \triangleright$ for group G at locator ID $\mathcal{K}$
1 **for** all $\mathcal{N}_i$ IDs in MFT_G
2 $\quad\quad$ **do if** $LCP(\mathcal{C}, \mathcal{N}_i) = \mathcal{C}$
$\quad\quad\quad\quad \triangleright \mathcal{N}_i$ is downtree neighbor
3 $\quad\quad\quad\quad$ **then** $\mathcal{C}_{new} \leftarrow \mathcal{N}_i$
4 $\quad\quad\quad\quad\quad\quad$ FORWARD PACKET TO $\mathcal{C}_{new}$

As the forwarding algorithm directly corresponds to a data dissemination along the multicast prefix tree, multicast coverage follows from the previous arguments. Furthermore, it is easy to show that all multicast listeners receive the data exactly once and the algorithm terminates.

2.2 Protocol Properties

The core protocol creates and manages a generic shared family of source trees in prefix space, which allow for unique multicast data transmission from any node in a prefix-optimized fashion. Assuming prefix aggregation to account for network topologies, replication will naturally follow locality and strictly retain provider bounds. This basic scheme is open to a variety of features as desired by the application or network scenario. In the following, we sketch such options, which come into operation without increasing the signaling load or management overhead.

Multipath Transport. All peers in a network that operates prefix-directed multicast are equally suited to serve as content distributors for a given group.

This capability may be exploited for resource pooling purposes [18] in the following way: Any *multi-homed* source of data streams may choose to distribute packets arbitrarily among its interfaces. Obtaining feedback from application level multicast transport (e.g., RTCP) will allow the sender to balance and load-shift traffic in response to current network conditions. Like in unicast, multipath selection will solely rely on end system behavior, while the multicast network offers a transparent, ubiquitous distribution layer.

Conversely, any *single-homed* source may utilize other peers for relaying, as any multicast forwarder may attain the role of a network relay. A multicast peer receiving a packet will continue distribution according to its destination prefix, which is equivalent of being the root of the multicast distribution sub-tree. Hence, a source can activate relays by simply sending packets with empty destination prefix. Multipath transport for single-homed peers is thus facilitated by chosing a first-hop relay. Furthermore, any intermediate peer can react on link failures by passing a packet with unaltered destination prefix to a life neighbor. The latter option on seamless routing resilience happens in contrast to BGP failure convergence times of 2 to 15 minutes [20].

r-Redundancy for Data and Paths. In the presence of network unreliability, it is desirable to add a certain degree of data redundancy to the distribution system. More precisely, an r-redundancy tolerating the loss of one packet out of each sequence of r may be sufficient to sustain information integrity at an appropriate degree of confidence. An efficient method to procure redundancy at the packet level has been recently introduced by network coding [21]. A straight application of this scheme to $r + 1$ subsequent packets leads to an r-redundancy for data.

Many disruptive scenarios like link and node failures are likely to cause damage at more than one occasional packet. Redundant paths are required to circumvent these defects. Relaying provides a natural option for path redundancy. A source willing to distribute data at an r-redundant level will perform network coding for r-sequences of its packet stream. It will further select r relays, preferably of differing prefix initials, as discussed in the previous section on multipath transport. Each sequence of r packets as well as the coded datagram are then distributed among the $r + 1$ senders. Choosing senders of different prefix initials will minimize a coincidence of the $r + 1$ prefix trees and will thus lead to the highest likelihood of completely disjoined paths. Any node/link failure occurring on a single distribution tree will then cause loss of one packet per r-sequence and can be fully compensated by the remaining r data units. Thus prefix-directed multicast can provide a full r-redundancy in data and paths without modification or additional signaling.

Mobility Support. A strong tendency can be observed to run infotainment group applications on mobile devices. The need consequently arises for a multicast mobility management, which in general turned out to be a complex problem [22]. The prefix-directed multicast leads to a mobility-agnostic routing environment in the sense that senders can seamlessly transmit multicast data from any location, while listeners may need to activate prefix branches for distribution,

which are in network proximity for regional moves. In this sense, our scheme represents a shared routing availability on inter-domain scale, similar as foreseen by BIDIR-PIM [23] for interior domains.

3   OASIS: Prefix-Directed Multicast on the IP(v6) Layer

In this section, we want to address a direct applicability of our solution to the Internet routing layer. It springs to mind that the current Classless Inter-domain Routing is based on prefix routing and aggregation. However, three major differences between the assumption of section 2.1 and the current Internet routing remain evident:

1. While BGP ensures complete visibility of prefixes, interior protocols provide knowledge only on a small subset.
2. A node may be responsible for multiple, non-adjacent prefixes.
3. The IPv4 address space is mainly unstructured, whereas attempts are taken to sustain a structure of prefix aggregation in IPv6.
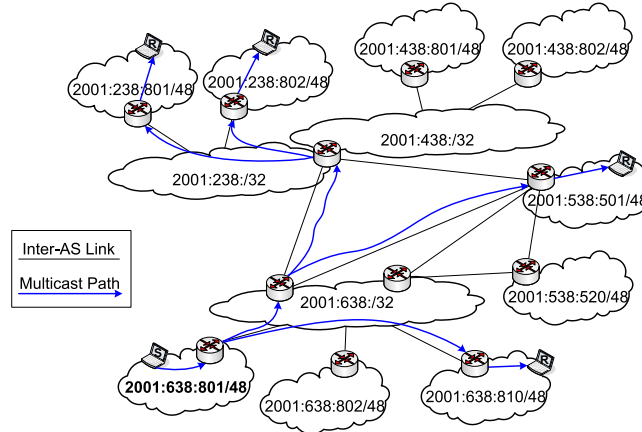
Thus, an immediate, unaltered transfer of our prefix-directed multicast distribution is not applicable, but strong correspondences in particular hold for IPv6.

Internet routers perform a prefix aggregation inherently, i.e., an upstream router can always keep track of the aggregation level it serves towards its downstream peers. Accordingly, a multicast join (or leave) received from a downstream interface can be agglomerated with parallel joins to be tied to the prefix level in operation. Consequently, any router in the Internet will be able to identify its own prefix aggregation level as well as paths to its prefix neighbors.

Assuming a clear, hierarchical address structure as proposed in RFC 2374 [24], group management on the BGP level could proceed as in our proposed scheme and direct message forwarding only downward. Instead of a single destination prefix, packets will carry multiple, non-adjacent prefixes as announced by the next hop peer in a dedicated extension header. Serving all prefixes of neighboring nodes simultaneously will avoid duplicate packet transmission.

Unfortunately, RFC 2374 has been obsoleted due to the persistent customer demand for provider independent addresses. To cope with unstructured address organization, a router receiving a join message for a specific multicast group needs to flood the join message under partial prefix aggregation to its remaining downstream *and* upstream interfaces, if the corresponding prefix has not been signaled before. Subsequent joins only need to forward up to a level of aggregation, where they are uncovered by previously joining group members. In this way, a shared prefix tree will be erected throughout the Internet, as visualized in figure 2.

The destination for a prefix will be selected with respect to the underlying unicast routing, which is optimal from the provider perspective. In the case of multi-homed sites, i.e., for a prefix assignment to several peers, one topological correct path will be chosen as in BGP.

Fig. 2. Multicast prefix tree spanning IPv6 members on inter-domain scale

After the tree construction, multicast routing can proceed on a bidirectional shared tree as described in section 2. Routers will be equipped with a shared virtual multicast forwarding table in prefix space, which gives rise to a forward route selection.

Prefix Mcast Forwarding on IPv6

    ▷ On arrival of packet with destination
    ▷ prefix list $\{\mathcal{C}\}$ for group G
1   **for** all $\{\mathcal{N}_i\}_R$ IDs in MFT_G grouped by next hop R
2   **do for** all $\mathcal{N}_i' \in \{\mathcal{N}_i\}_R$ with $LCP(\mathcal{N}_i', \mathcal{C}) = \mathcal{C}$, $\mathcal{C} \in \{\mathcal{C}\}$
3   **do** $\{\mathcal{C}_{new}\} \cup \{\mathcal{N}_i'\}$
4   Forward packet to $\{\mathcal{C}_{new}\}$

Prefix-directed multicast distribution transparently covers inter- and intra-domain routing. No supplementary mechanisms to implement inter-domain group transition such as MSDP are needed.

It may be undesirable to initiate an Internet-wide distribution tree for all multicast addresses. To enable scoping of restricted prefix ranges in multicast dissemination, one could proceed as follows: On creation of the group, a uniformly covering prefix[1] can be assigned and embedded into the multicast group address in analogy to a rendezvous point (RP) address [25]. This prefix would then be interpreted as the root of the prefix tree, preventing state distribution beyond the region of interest for the group in common. It is worth noting, that the root is virtual and the embedded prefix does not predefine a corresponding forwarder in contrast to the embedded RP address. The efficiency of multicast forwarding is not affected.

---

[1] The prefix 2001:0638::/32 could for example be chosen to restrict a group to the German NREN community.

Within a cleanly aggregated address space, the prefix routing described above will be strictly bound to provider borders and – if required – will cross peering links exactly once. All fundamental properties, especially redundancy and mobility options, are inherited from prefix-based multicast as explained in section 2.

The IPv6 address allocation and assignment policy of APNIC, ARIN and RIPE NCC [26] mostly comply with the IAB/IESG recommendations on IPv6 address allocations to sites [27]. Normally, prefixes of length 32 will be assigned to providers and prefixes of length 48 to end sites. A current BGP table snapshot of the RouteView project reflects this scheme: Approximately 65% of the announced prefixes exhibit a length of 32 bits and 20% 48 bits. Provider independent (PI) addresses are under discussion [28] or already been scheduled [29] by the RIRs. The RIPE policy proposal suggests PI assignments allocated from a dedicated address block, which allows to identify and subsume PI addresses. Furthermore, address indirection approaches like LISP [30], or Six/One [31] try to regain provider-bound address hierarchies by splitting locators and identifiers.

4   Analytical Evaluation

The well defined prefix structure of the multicast scheme allows for a detailed theoretical analysis, yielding strong analytical results for all major properties. For the sake of brevity, we will enumerate only the cost aspects of group management, cf. [19] for proofs and further results.

For a given key space of alphabet size k, we consider the corresponding k-ary prefix tree as basic structure. Therein N overlay nodes $\{\mathcal{N}\}$ are uniformly placed at leaf nodes of the prefix tree. The prefix-directed algorithm aggregates multicast receivers. For a given prefix $\mathcal{C}$ of length j, the probability that a receiver shares $\mathcal{C}$ is therefore of general relevance.

Theorem 1. *For a multicast group G, a k-ary alphabet and N nodes, the probability that a prefix $\mathcal{C}$ of length j is attained by at least one of g receivers reads*
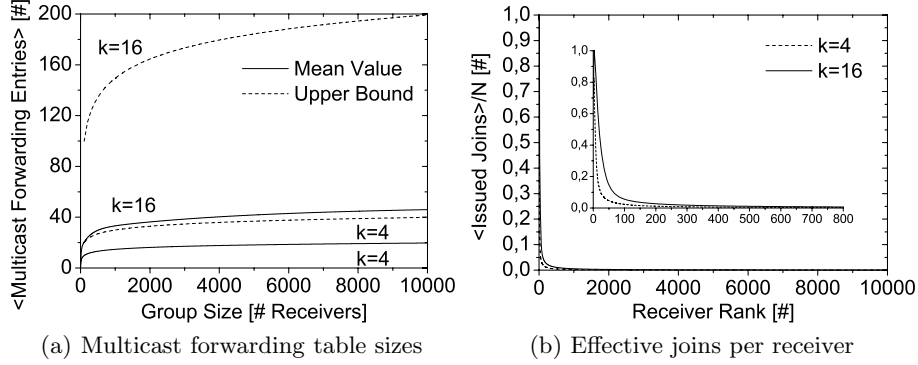
$$P\left(\left|\left\{\mathcal{G} \in G\mid LCP(\mathcal{C}, \mathcal{G}) = \mathcal{C}\right\}\right| \geq 1\right)$$
$$= 1 - \left(1 - \frac{g}{k^j N}\right)^N = 1 - e^{-\frac{g}{k^j}} + \mathcal{O}\left(\frac{1}{N}\right). \tag{1}$$

It is worth noting that in large networks the prefix distribution of multicast receivers is effectively independent of the network size.

From theorem 1 we can calculate the *size of multicast forwarding tables:*

Theorem 2. *For any peer in a k-ary prefix tree with g leaf nodes (receivers), the number of adjacent vertices is limited by $\log_2(g)(k-1)$. This bound equally limits the number of multicast forwarding table entries. Furthermore, the probability distribution $P(j, l)$ that a given overlay node holds l multicast forwarding entries at prefix level j reads*

$$P(j, l) = \binom{k-1}{l}\left(1 - e^{-\frac{g}{k^{j+1}}}\right)^l \left(e^{-\frac{g}{k^{j+1}}}\right)^{k-1-l} + \mathcal{O}\left(\frac{1}{N}\right).$$

(a) Multicast forwarding table sizes   (b) Effective joins per receiver

Fig. 3. Analytical results for group management costs

Mean values and upper bounds are plotted in figure 3(a). Table entries remain significantly below its upper bound, reproducing nicely the logarithmic dependency on g. In the remaining we quantify the *signaling load*. Due to the symmetry of operations, the following analysis is restricted to the join procedure.

Theorem 3. *The probability $P(j,g)$ for distributing a* JOIN *or* LEAVE *message within a prefix tree at injection level j reads*

$$P(j,g) = \left(1 - e^{-\frac{g}{k^j}}\right) e^{-\frac{g}{k^{j+1}}} + \mathcal{O}\left(\frac{1}{N}\right), \qquad (2)$$

where g is the number of group members prior to signaling. Moreover, the expected ratio of flooded nodes is well approximated by

$$(1-e^{-g})e^{-\frac{g}{k}} + \frac{k}{g(k+1)\ln k}\left(\left(e^{-\frac{g}{k^{h+1}}} - e^{-\frac{g}{k}}\right)(k+1) + e^{-\frac{g(k+1)}{k}} - e^{-\frac{g}{k^{h+1}}(k+1)}\right).$$

The results are displayed in figure 3(b) as functions of the joining receiver rank, where the insert provides a detail zoom. Signaling expenses admit a strong exponential decay in the expected number of flooded nodes. The mean number of messages issued for Join/Leave reduces to below 1 % for group sizes above 500, keeping group management costs marginal in larger distribution settings.

5   Conclusions and Outlook

Starting from the generic principles of naming, unicast routing and locator aggregation, we proposed and analyzed a forward-path multicast forwarding mechanism directed by an overlay abstraction in aggregation space. Combining structured overlay algorithms with Internet routing, data is guided along logarithmically scalable source specific shortest path trees, even though the scheme

complies with the general Any Source Multicast semantic. Without further signaling, the protocol facilitates multipath transport for load sharing and redundancy, seamless mobility support, as well as an inherently transparent transition between the underlay and a corresponding overlay. The latter aspect is of particular interest for an incremental deployment, as it allows for an easy integration of the multicast routing in hybrid architectures, which may span agnostic regions by overlay transits.

In future work, we will elaborate on further performance evaluations and deployment concepts, targeting on different mappings of the prefix space to underlay, overlay and hybrid network scenarios.

Acknowledgements

References

1. Aguilar, L.: Datagram Routing for Internet Multicasting. In: Proceedings of SIG-COMM 1984, pp. 58–63. ACM Press, New York (1984)
2. Deering, S.E.: Host Extensions for IP Multicasting. RFC 1112, IETF (1989)
3. Castro, M., Druschel, P., Kermarrec, A.M., Rowstron, A.: SCRIBE: A large-scale and decentralized application-level multicast infrastructure. IEEE Journal on Selected Areas in Communications 20(8), 100–110 (2002)
4. Zhuang, S.Q., Zhao, B.Y., Joseph, A.D., Katz, R.H., Kubiatowicz, J.D.: Bayeux: An Architecture for Scalable and Fault-tolerant Wide-Area Data Dissemination. In: Proceedings of NOSSDAV 2001, pp. 11–20 (June 2001)
5. Ratnasamy, S., Handley, M., Karp, R.M., Shenker, S.: Application-Level Multicast Using Content-Addressable Networks. In: Crowcroft, J., Hofmann, M. (eds.) NGC 2001. LNCS, vol. 2233, pp. 14–29. Springer, Heidelberg (2001)
6. Thaler, D.: Evolution of the IP Model. Internet Draft – work in progress 01, IETF (July 2008)
7. Saltzer, J.H., Reed, D.P., Clark, D.D.: End-to-End Arguments in System Design. ACM Trans. Comput. Syst. 2(4), 277–288 (1984)
8. Clark, D.: The Design Philosophy of the DARPA Internet Protocols. In: Proceedings of SIGCOMM 1988, pp. 106–114. ACM, New York (1988)
9. Carpenter, B.: Architectural Principles of the Internet. RFC 1958, IETF (1996)
10. Saltzer, J.: On the Naming and Binding of Network Destinations. RFC 1498, IETF (August 1993)
11. Braden, R., Clark, D., Shenker, S., Wroclawski, J.: Developing a Next-Generation Internet Architecture. White paper (July 2000)
12. Crowcroft, J., Hand, S., Mortier, R., Roscoe, T., Warfield, A.: Plutarch: an Argument for Network Pluralism. In: Proc. of the ACM SIGCOMM workshop on Future directions in network architecture, pp. 258–266. ACM Press, New York (2003)

13. Schmid, S., Eggert, L., Brunner, M., Quittek, J.: TurfNet: An Architecture for Dynamically Composable Networks. In: Smirnov, M. (ed.) WAC 2004. LNCS, vol. 3457, pp. 94–114. Springer, Heidelberg (2005)
14. Plaxton, C., Rajaraman, R., Richa, A.: Accessing Nearby Copies of Replicated Objects in a Distributed Environment. In: Proc. of 9th ACM Sympos. on parallel Algor. and Arch. (SPAA), pp. 311–330. ACM Press, New York (1997)
15. Stoica, I., Morris, R., Karger, D., Kaashoek, M.F., Balakrishnan, H.: Chord: A scalable peer-to-peer lookup service for internet applications. In: Proceedings of SIGCOMM 2001, pp. 149–160. ACM Press, New York (2001)
16. Rowstron, A., Druschel, P.: Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In: Guerraoui, R. (ed.) Middleware 2001. LNCS, vol. 2218, pp. 329–350. Springer, Heidelberg (2001)
17. Ratnasamy, S., Francis, P., Handley, M., Karp, R., Schenker, S.: A Scalable Content-Addressable Network. In: Proc. of SIGCOMM 2001, pp. 161–172. ACM, New York (2001)
18. Wischik, D., Handley, M., Braun, M.B.: The Resource Pooling Principle. SIGCOMM Comput. Commun. Rev. 38(5), 47–52 (2008)
19. Wählisch, M.: Scalable Adaptive Group Communication on Bi-directional Shared Prefix Trees. Technical Report TR-B-08-14, Freie Universität Berlin, Department of Mathematics and Computer Science, Berlin (September 2008)
20. Labovitz, C., Ahuja, A., Bose, A., Jahanian, F.: Delayed Internet Routing Convergence. In: Proceedings of SIGCOMM 2000, pp. 175–187. ACM, New York (2000)
21. Ahlswede, R., Cai, N., Li, S.Y.R., Yeung, R.W.: Network Information Flow. IEEE Transactions on Information Theory 46(4), 1204–1216 (2000)
22. Schmidt, T.C., Wählisch, M., Fairhurst, G.: Multicast Mobility in MIPv6: Problem Statement and Brief Survey. IRTF Internet Draft – work in progress 07 (2009)
23. Handley, M., Kouvelas, I., Speakman, T., Vicisano, L.: Bidirectional Protocol Independent Multicast (BIDIR-PIM). RFC 5015, IETF (October 2007)
24. Hinden, R.M., O'Dell, M., Deering, S.E.: An IPv6 Aggregatable Global Unicast Address Format. RFC 2374, IETF (July 1998)
25. Savola, P., Haberman, B.: Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address. RFC 3956, IETF (November 2004)
26. APNIC, ARIN, RIPE NCC: IPv6 Address Allocation and Assignment Policy. RIPE Document ripe-421, RIPE (November 2007)
27. IAB, IESG: IAB/IESG Recommendations on IPv6 Address Allocations to Sites. RFC 3177, IETF (September 2001)
28. Martinez, J.P.: Provider Independent (PI) IPv6 Assignments for End User Organisations. RIPE Policy Proposal 2006-01, RIPE (May 2007)
29. APNIC: IPv6 Address Allocation and Assignment Policy. APNIC Document APNIC-089, APNIC (August 2008)
30. Farinacci, D., Fuller, V., Oran, D., Meyer, D.: Locator/ID Separation Protocol (LISP). Internet Draft – work in progress 09, IETF (October 2008)
31. Vogt, C.: Six/One: A Solution for Routing and Addressing in IPv6. Internet Draft – work in progress (expired) 01, IETF (November 2007)