Scalability Analysis of Source Routing Multicast for Huge Numbers of Groups

Yohei KATAYAMA†, Takeru INOUE‡*, Noriyuki TAKAHASHI†, and Ryutaro KAWAMURA†, Members

SUMMARY Source routing multicast has been gathering much more attention rather than traditional IP multicast, since it is thought to be more scalable in terms of the number of groups at the cost of higher traffic loads. This paper introduces a mathematical framework to analyze the scalability of source routing multicast and IP multicast by leveraging previous multicast studies. We first analyze the amount of data traffic based on the small-world nature of networks, and show that source routing multicast can be as efficient as IP multicast if a simple header fragmentation technique (subgrouping) is utilized. We also analyze scalability in terms of group numbers, which are derived under the equal budget assumption. Our analysis shows that source routing multicast is competitive for low bit-rate streams, like those in the publish/subscribe service, but we find some factors that offset the advantage. This is the first work to analytically investigate the scalability of source routing multicast.

key words: multicast, source routing, stateless, publish/subscribe

1. Introduction

Multicast is a technique used for efficient information delivery. Since a single packet emitted by the data source is replicated by intermediate routers for delivery to a group of potential many recipients, source overhead is reduced and overall network traffic is decreased. In the last few decades, enormous effort has been made to develop several multicast schemes including IP multicast (IPM) [1]. IPM is being used for content delivery to just small numbers of groups, because it does not scale well against group number; this is because routers are required to store forwarding state for each group to maintain forwarding paths. Multicast networks, however, will be required to accommodate many more groups [2]–[5], because of the great increase in the number of contents and the significant growth in long-lived services that follow the publish/subscribe paradigm [6]. Each data source in publish/subscribe produces small intermittent traffic, but the total traffic volume can become huge due to the great number of groups; this traffic volume should be well handled by multicast [7].

The research community is seeking a novel multicast scheme that can handle the large number of groups expected [8]. To the best of our knowledge, there are two main schemes. One scheme, shown in [2], [9]–[12], extends IPM to improve scalability by minimizing the forwarding state. The other, shown in [4], [5], [7], [13], [14], puts the multicast forwarding path into the packet’s header, instead of relying on the routers to hold the forwarding state. The latter, which is often called source routing multicast (SRM), is believed to handle more groups than IPM, since routers are not required to maintain forwarding state. SRM was previously considered to be impractical due to its large header; the assumption was that the header would have to hold IP addresses of all routers on the multicast forwarding path. However, free riding multicast (FRM) [14] offers a novel way to put the multicast forwarding path in the header; it utilizes Bloom filters [15] so each link occupies only around 10 bits on the header, regardless of the original length of the link identifiers.

Researchers have been steadily improving scalability but no one has, up to now, produced a sound comparison between the two multicast schemes. This paper introduces a mathematical framework to compare the two multicast schemes by leveraging previous multicast research [14], [16]–[20]. It reveals which scheme has greater scalability. Our contributions are as follows.

• Traffic analysis. We first analyze the data traffic created by the two multicast schemes. In the analysis, we take advantage of the small-world nature of underlying networks and derive the asymptotic behavior of the traffic. The results show that SRM can yield much heavier traffic than IPM, but that the use of subgrouping makes SRM as efficient as IPM. Subgrouping is a traditional technique in which a large header is cut into small fragments [14], [17]. Its efficiency has not, up to now, been examined analytically.

• Scalability analysis. We analyze scalability in terms of the maximum number of groups accommodated in a network. The number is evaluated under the equal budget constraint. The results show that SRM is competitive for low bit-rate streams (e.g. publish/subscribe service). We also find the condition under which the competitiveness of SRM degrades significantly. Unexpectedly, group size, which determines the header size, is not a significant factor due to subgrouping.

In our scalability analysis, we focus on resource consumption by the forwarding process of data traffic, not by the routing process (the latter includes management of multicast address, group membership and network topology). This is because the forwarding process directly relates to performance and suffers more from resource constraints, such as expensive forwarding memories and expensive switching modules.
necessary for wire rate, than the routing process does, which relates to feasibility and deployability. Of course, routing is an important issue and feasibility is discussed in [7], [14].

Our analysis focuses on clarity rather than details because our aim is to determine which multicast scheme has greater scalability, not to precisely calculate the maximum number of groups in a network. Likewise, we take the average behavior of groups into consideration, and ignore the differences between groups. Our analysis is not dependent on any particular deployment scenario like inter- or intra-domain multicast, since we tackle the fundamental problems of multicasting. We believe that our work gives a good starting point for assessing the two multicast schemes and opens the way to more precise analyses.

The rest of this paper is organized as follows. Section 2 describes the two multicast schemes. Section 3 examines the traffic created by the two schemes, and Sect. 4 compares their scalabilities. Section 5 summarizes related work. We finally conclude this paper in Sect. 6.

2. Two Multicast Schemes

We explain the terminology of multicast using Fig. 1, before explicating the two multicast schemes. Without loss of generality, we discuss only groups that have a single source as shown in the figure. A router connected to a data source is called a root router, while those connected to recipients are called destination routers. The tree-like forwarding path from a root router to recipients is referred to as the delivery tree.

2.1 IP Multicast

We illustrate an example of the packet forwarding mechanisms of IPM in Fig. 1(a). Every router holds a forwarding table in memory, as shown in the figure. The table is populated with forwarding entries of groups that the router serves. Each entry consists of the group address and corresponding output ports (links); e.g., “G1: L1, L2” indicates an forwarding entry where the key is group address G1 associated with output links L1 and L2. Upon receiving a packet, the router searches its forwarding table using the group address contained in the packet header, and determines the output ports if a match is found; e.g., in the figure, the root router forwards the packet of G1 to links L1 and L2 following its forwarding table.

Figure 2(a) shows an example of router architecture of IPM. As shown in the figure, every group requires a forwarding entry in the linecard memory, which is often fast and expensive memory (e.g., SRAM) for fast table lookup. In addition, every packet goes through the switch fabric, and the bandwidth is consumed depending on the packet rate. In this paper, we assume that the maximum number of groups handled by an IPM router is limited by either memory size for forwarding tables or switching bandwidth of the fabric.

We focus on the original IPM in the analysis, and discuss its extensions in Sect. 5.

2.2 Source Routing Multicast

Figure 1(b) provides the packet forwarding mechanisms of SRM. When a root router receives a data packet from a source, the router puts the delivery tree of the group into the packet header1; e.g., in the figure, the tree consists of links L1–4 and L6–9. Upon receiving the packet, intermediate routers find their own links in the header and output the packet to these links; e.g., the root router sends the packet into L1 and L2, which are found on the header. The delivery tree is constructed every time a packet is sent, or is constructed and cached in a preliminary step to suppress repetitive computation [14] (we ignore memory required for this cache for a while, and the impact of the delivery tree cache will be discussed in Sect. 4.3). A root router (or another computation element, such as a source) maintains a network map and group membership, in order to calculate the delivery trees.

We show SRM router architecture in Fig. 2(b). Routers maintain no forwarding table unlike IPM, but they are re-

1In Fig. 1(b), the whole delivery tree is put into the header of every packet that traverses link L2, even though the packet will not pass through the left-side links (L1, L3, L4, ...). This inefficiency is resolved in Sect. 3.3 and is ignored here to simplify the explanation.
quired to handle large-header packets, which pass through the switch fabric as shown in the figure. We assume that the maximum number of groups handled by an SRM router is determined only by the switching bandwidth of the fabric because a delivery tree in the packet header consumes switching bandwidth but not memory. We also assume that the delivery tree on the header is unchanged during packet forwarding for the purpose of fast packet processing.

3. Traffic Analysis

This section examines the amount of traffic needed for delivery to a single group. Our methodology in this section is to derive the asymptotic behavior of the traffic, which is then compared with calculations. The behavior is described against group size, which is the number of recipients in the group and has a major impact on the traffic. We discuss the average traffic among groups, since details of individual groups are not our concern.

We review the representative delivery tree, which was introduced in [16] and provides the basis of our analysis, in Sect. 3.1. We then analyze the traffic of the two multicast schemes in Sect. 3.2 and examine the impact of subgrouping [14], [17], which reduces the traffic of SRM, in Sect. 3.3.

Parameters and values used in this paper are presented in Tables 1 and 2.

### 3.1 Representative Delivery Tree

Traditionally, IP multicast is compared with unicast in terms of traffic or bandwidth consumption, the number of traversed links needed to reach destination routers from a root router, $l$ (for example, $l = 8$ in Fig. 1). References [16], [18] found that the average value of $l$'s can be well estimated by introducing a model tree built on an exponentially growing network like the Internet. Its accuracy was confirmed in [19], [20]. We call this model tree the representative delivery tree, and use it to analyze the multicast traffic instead of extracting many sample trees from a network.

Before mathematically defining the representative delivery tree, we clarify the four key assumptions made in applying it.

---

**Table 1** Summary of main parameters.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>degree of exponential growth of the network</td>
</tr>
<tr>
<td>$R$</td>
<td>average distance between routers or network radius</td>
</tr>
<tr>
<td>$n$</td>
<td>group size</td>
</tr>
<tr>
<td>$f$</td>
<td>network-wide traffic amount for delivery to a group</td>
</tr>
<tr>
<td>$l$</td>
<td># of traversed links to reach destinations</td>
</tr>
<tr>
<td>$\lambda$</td>
<td># of routers on a delivery tree</td>
</tr>
<tr>
<td>$h$</td>
<td>header size</td>
</tr>
<tr>
<td>$b$</td>
<td># of bits required to add one link of a delivery tree to a header (SRM only)</td>
</tr>
<tr>
<td>$p$</td>
<td>payload size</td>
</tr>
<tr>
<td>$d$</td>
<td># of subgroups (SRM only)</td>
</tr>
<tr>
<td>$r$</td>
<td>packet rate</td>
</tr>
<tr>
<td>$s$</td>
<td>forwarding entry size (IPM only)</td>
</tr>
<tr>
<td>$m$</td>
<td>memory size (IPM only)</td>
</tr>
<tr>
<td>$w$</td>
<td>switching bandwidth</td>
</tr>
<tr>
<td>$g$</td>
<td>max # of groups at a router</td>
</tr>
<tr>
<td>$G$</td>
<td>max # of groups in the network</td>
</tr>
<tr>
<td>$c$</td>
<td>router price</td>
</tr>
<tr>
<td>$C$</td>
<td>budget for routers</td>
</tr>
<tr>
<td>$N$</td>
<td># of routers in the network</td>
</tr>
</tbody>
</table>

**Table 2** Default values (used as stated unless specified otherwise).

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>3</td>
<td>measured in Sprint network [21]</td>
</tr>
<tr>
<td>$R$</td>
<td>6</td>
<td>measured in Sprint network [21]</td>
</tr>
<tr>
<td>$n$</td>
<td>100</td>
<td>small group size</td>
</tr>
<tr>
<td>$h_I$</td>
<td>128 bits</td>
<td>IPv6 address length</td>
</tr>
<tr>
<td>$b$</td>
<td>16 bits</td>
<td>conservative compared with [13], [14] (about 10 bits)</td>
</tr>
<tr>
<td>$p$</td>
<td>8 Kbits</td>
<td>roughly equal to or smaller than that of publish/subscribe applications like Twitter or RSS.</td>
</tr>
<tr>
<td>$s$</td>
<td>128 bits</td>
<td>IPv6 address length</td>
</tr>
<tr>
<td>$r$</td>
<td>1pps</td>
<td>just for tentative plotting</td>
</tr>
</tbody>
</table>

---

"We ignore the memory space required to maintain the neighboring links at each router because it is a negligible amount [14]; it is not significant in comparing the two multicast schemes."
• Delivery trees are constructed on exponentially growing networks. The exponential growth characteristic is commonly seen in the so-called small world networks [22] including the Internet.
• All links are bidirectional and have the same link weight.
• A delivery tree consists of shortest paths from a root router to destination routers, as most of the multicast routing protocols follow this assumption to reduce computation complexity [23], [24].
• The recipients are uniformly randomly located at leaf sites of the delivery tree simply since reference [16] showed that the asymptotic behavior of \( l \) was unchanged when the recipients were located at sites of the delivery tree.

We briefly review the asymptotic form of the delivery tree. In exponentially growing networks, the number of routers (and links) that are \( i \) hops apart from a certain router increases exponentially with \( i \); i.e., there are, on average, \( k^i \) routers that are \( i \) hops from a router, where \( k \) is a constant. This means that a shortest path tree rooted at a router can be modeled by a \( k \)-ary tree of depth \( R \), where \( R \) is average distance between routers or network radius. Figures 1(a) and (b) show examples with \( k = 2 \) and \( R = 3 \).

We put \( n \) recipients at randomly chosen leaves of the \( k \)-ary tree, and count the number of links on the path. Since the number of on-path links that are \( i \) hops apart from the root is \( k^i \cdot (1 - (1 - \frac{1}{k^i})^n) \), the number of links traversed to reach the destination routers is, on average, given by,

\[
l = \sum_{i=1}^{R} k^i \left( 1 - \left( 1 - \frac{1}{k^i} \right)^n \right).
\]

Likewise, the number of routers on a delivery tree is, on average, given by,

\[
\lambda = \sum_{i=0}^{R-1} k^i \left( 1 - \left( 1 - \frac{1}{k^i} \right)^n \right).
\]

References [18] found the scaling law of \( l \) as follows,

\[
l = O(n^{\zeta}),
\]

where \( \zeta \) is a constant determined by \( k \) and \( R \), and references [18], [25] found \( \zeta \approx 0.7 \) or 0.8 with such sets of \( k \) and \( R \) as appear in Internet-like topologies, while the unicast traffic increases in proportion to the number of recipients, \( nR = O(n) \). This asymptotic behavior provides a measure of the traffic reduction yielded by multicast.

3.2 Measurement of Traffic

The traditional analysis approach in the previous subsection focuses on \( l \) and ignores packet size because packet size is of little consequence in a traffic comparison involving IPM and unicast. However, packet size plays a key role in traffic comparisons between IPM and SRM. We measure network-wide traffic, \( f \), as follows,

\[ f = lr(h + p), \]

where \( r \) is packet rate at the source, and \( h \) and \( p \) are packet header size and payload size, respectively. In this paper, the word header is redefined to include only forwarding information on the header (e.g., a group address and/or a delivery tree), while payload includes all other fields of the packet.

The header size in IPM, \( h_I \) (the subscript \( I \) means IPM), represents just a group address. For the header size in SRM, we make two assumptions. First, we assume that the header includes a group address because a source or recipients may use a group address to distinguish between groups. Second, we assume that the header grows in proportion to tree size \( l \). Thus the header size in SRM, \( h_S \) (the subscript \( S \) means SRM), is given by,

\[ h_S = bl + h_I, \]

where \( b \) is the number of bits required to add one link of a delivery tree to a header.

We obtain the traffic amounts for IPM and SRM, as follows.

\[ f_I = lr(h_I + p) = O(l) < O(n), \] (3)
\[ f_S = lr(h_S + p) = bl^2r + lr(h_I + p) = O(l^2) > O(n). \] (4)

The inequality in (4) means that SRM produces more traffic than unicast at large group sizes. We plot \( f_I \) and \( f_S \) in Fig. 3 (the parameters \( k = 10 \) and \( R = 4 \) are taken from an Autonomous System (AS) level Internet topology measurement [26]). The additional lines in the figure, \( O(n^{1.6}) \) and \( O(n^{0.8}) \), reinforce the asymptotic behaviors of the traffic in IPM and SRM.

3.3 Subgrouping on Source Routing Multicast

We analyze the effect of subgrouping [14], [17]. Subgrouping means to divide a large group into small subgroups to shrink the large headers possible in SRM. Subgrouping, however, increases traffic around root routers, because they are required to transmit as many packets as there are subgroups. To minimize this inefficiency, we assume that as many recipients on the same subtree belong to same subgroup as possible, as shown in Fig. 4. The header size of

![Fig. 3 Average traffic needed for delivery from a root router to destination routers (log-log plot).](image-url)
each packet is reduced by subgrouping, while the cumulative number of links on all the subgroup trees can be greater than that of the original group (e.g., $l_D = 24 > l = 22$, as redundant packets would pass through the upper links; the subscript $D$ means SRM with subgroups).

We now investigate whether subgrouping reduces the total traffic; we examine the number of total links needed to reach all subgroups, $l_D$, and the header size with subgrouping, $h_D$. Parameter $d$ is the number of subgroups, where each subgroup has $n/d$ recipients on average. For a single subgroup, the number of links that are $i$ hops apart from the root is $\lceil k'/d \rceil$ ($\lceil x \rceil$ is a smallest integer not less than $x$). We, therefore, obtain the number of links to reach a single subgroup, $l_D'$, as follows,

$$l_D' = \sum_{i=1}^{R} \lceil k'/d \rceil \left( 1 - \frac{1}{\lceil k'/d \rceil} \right)^{n/d},$$  \hfill (5)

where we replace $n$ with $n/d$ and $k'$ with $\lceil k'/d \rceil$ in (1). The number of links to reach all subgroups, $l_D$, is simply given by,

$$l_D = dl_D'.$$

As discussed in Appendix A, subgrouping yields relatively fewer redundant packets if the original group is large enough and only a few subgroups are used. Thus we have the following approximation for $n \gg 1$ and $d \ll n$,

$$l_D \approx l + h_D.$$  \hfill (6)

With (6), the header size with subgrouping, $h_D$, is given by,

$$h_D = bl_D' + h_I \approx bl/d + h_I.$$  \hfill (7)

Finally, the traffic with subgrouping, $f_D$, is given by,

$$f_D = l_D r (h_D + p) \approx lr \left( \frac{bl}{d} + h_I + p \right).$$

This expression indicates that the traffic decreases with $d$. We confirm this behavior by the plot of $f_D$ in Fig. 5, which is numerically calculated without approximation. The figure shows that the traffic is reduced by subgrouping for small numbers of subgroups ($d \ll n$). For larger $d$, it is hard to understand the traffic behavior analytically due to the heavy fluctuation of the ceiling functions in (5). We here discuss this behavior qualitatively, which shows the traffic is likely to increase for larger $d$ as follows. When dividing groups (i.e., increasing $d$), smaller headers of larger $d$ are less reduced than larger ones of smaller $d$. Moreover, more redundant packets (more links) are added with more delivery trees of larger $d$.

We next examine the asymptotic behavior of the traffic with subgrouping. We could not derive an analytical form of the optimum subgroup number, $d^*$, but the asymptotic behavior can be understood as follows. Since the header size at the optimum subgroup number $d^*$ is nearly constant with $n$ (we discuss this in Appendix B), we find $h_{D^*} = O(1)$. With this property, we find that the traffic follows $O(l)$ as follows,

$$f_D = l_D' r (h_{D^*} + p) \approx lr (O(1) + p) = O(l).$$

This asymptotic behavior indicates that subgrouping can reduce the traffic as much as IPM if constant factors are ignored. We plot the traffic of subgrouping in Fig. 3; $d^*$ is numerically discovered. The additional line in the figure, $O(n^{0.8})$, reinforces the asymptotic behavior of the traffic in SRM with subgroups. The constant factors are insignificant here.

As described above, we assume that recipients on the same subtree belong to same subgroup. This assumption is required to minimize the subgrouping traffic of (5), as follows. Without this assumption, the number of links $i$ hops apart from the root can be greater than $\lceil k'/d \rceil$; e.g., in Fig. 4, if recipients of G1-1 are placed on the left-most subtree as well as the second left subtree, the delivery tree of G1-1 has to branch the upper level and more links are required for delivery. This greater number of links means a larger tree as well as larger header, which yields more traffic.
From root to dest.s

- We briefly investigate the traffic generated by a root router. Since it is hard to derive the root traffic analytically, we first plot the numerical results in Fig. 6. The traffic growth roughly follows $O(l)$. We next present a sketch to understand the behavior. Subgrouping requires the root router to send out $d^r$ packets of size $h_D + p$, and the root traffic is roughly given by,

$$d^r(h_D + p) = blr \left( 1 + \frac{h_D + p}{h_D - h_I} \right) = O(l), \quad (8)$$

where we make use of (6) and (7), since we numerically confirmed that $d^r \ll n$ roughly hold in Fig. 5 and some other cases. Figure 6 and (8) indicate that the root traffic with subgrouping is much smaller than that of unicast, $O(n)$, though larger than that of IPM, $O(1)$, at large group sizes.

We summarize the results of the traffic analysis in Table 3.

### 4. Scalability Analysis

This section analyzes the scalability of the two multicast schemes, IPM and SRM with subgroups. Scalability is measured by the maximum number of groups that can be accommodated in a network. We begin with the maximum number of groups accommodated in a single router in Sect. 4.1, and then investigate the maximum number of groups accommodated in a network under a budget constraint in Sect. 4.2. To determine which multicast scheme has the greater scalability, we examine the ratio of the maximum number of groups between the two multicast schemes in the analysis. Section 4.3 derives an approximate form of the ratio, which helps us understand the key factors that determine which multicast scheme has greater scalability.

This section presents results for just one topology of an AS network, the Sprint network ($k = 3$ and $R = 6$), due to space limits. Our conclusions, however, still hold for other AS networks. This section considers multiple groups, while the previous section focused on a single group. We assume that all groups in the network have the same size $n$ and the same data rate $rp$ (the assumption of same $n$ is not critical as we will show later).

#### 4.1 The Maximum Group Number per Router

We analyze the maximum number of groups that a single router can handle, $g$. Here, we introduce a simple router model; the router has memory of size $m$ and switching bandwidth $w$ ($m$ is the size of memories needed for the forwarding tables in all of line cards or the switch fabric on the router, and $w$ is the minimum of the backplane bandwidth or all-port bandwidth just for output).

We derive values for IPM and SRM.

- **IPM.** The maximum group number on an IPM router, $g_I$, is determined by either $m$ or $w$ as discussed in Sect. 2.1. Memory size $m$ restricts $g_I$ as $m/s$, where $s$ is the size of a forwarding entry. We ignore the size of output port identifiers in an entry for simplicity (this simplification slightly overestimates the maximum group number on an IPM router). Switching bandwidth $w$ restricts $g_I$ as $w/\lambda$, where $f_I/\lambda$ is average output bit-rate per router for a single group (this is given by dividing total traffic to a group, $f_I$, by the number of routers on a delivery tree, $\lambda$). Thus, the maximum group number $g_I$ is given by,

$$g_I = \min \left( \frac{m}{s}, \frac{w}{f_I/\lambda} \right)$$

- **SRM.** Since the maximum group number on an SRM router, $g_D$, is limited only by switching bandwidth, it is given by,

$$g_D = \frac{w}{f_D/\lambda}$$

Since we do not consider flow alignment among ports in a router, we may overestimate the maximum group numbers by a constant factor against the number of ports in a router. However, this deviation is negligible because we can expect a statistical effect among massive numbers of groups.

We plot $g_I$ and $g_D$ versus data rate $rp$ in Fig. 7. The memory limit degrades $g_I$ at small data rates (e.g. publish/subscribe service), while large packet headers degrade $g_D$ overall. In publish/subscribe, average data rate is usually less than 1 kbps, as observed in Twitter and RSS. At this data rate level, an SRM router can accommodate more groups than an IPM router. Though publish/subscribe traffic is quite spiky, we consider just the average rate because we can expect a statistical effect among massive numbers of groups.

#### 4.2 The Maximum Group Number in a Network

We discuss the maximum group number in a network built...
Fig. 7  The maximum group number of IPM and SRM routers (we set $m = 9.2$ Gbits and $w = 3.4$ Tbps tentatively).

with multicast routers introduced in the previous subsection. We assume that all routers have the same configuration ($m$ and $w$). Let $N$ be the number of routers in the network and let $\lambda$ be the number of routers in a delivery tree; the maximum group number in a network, $G$, is given by,

$$G = N \frac{\lambda}{3}.$$  \hspace{1cm} (9)

It is worth noting that actual networks cannot accommodate $G$ groups, because delivery trees are non-uniformly allocated among routers and some routers become full before others. However, given that both multicast schemes will experience the same degree of non-uniformity, we expect that the ratio of two multicast scheme $G$’s remains valid.

We, next, introduce the budget for routers, $C$, to determine the optimum router configuration, i.e. the one that maximizes $G$, and the optimum router number in the network. The constraint is given by,

$$Nc \leq C,$$  \hspace{1cm} (10)

where $c$ is the price of a single router, $c_M$ and $c_W$ are prices of router memory and switching components (technically, the router price components of $m$ and $w$), respectively, and $c_0$ includes other costs. We do not consider other network equipment such as L2 switches and cables. We approximate $c_M$ and $c_W$ by cubic functions as follows. In most computing hardware, the price per capacity is a convex function of the capacity [27]. Assuming that the convex function is quadratic, we find that the price follows a cubic function by multiplying the quadratic function by the capacity. As shown in Fig. 8 and Fig. 9, market prices [28], [29] well fit to the cubic function curves. We set $c_0 = $250,000 [29]. The prices can vary depending on router vendors and purchase volume, but we have confirmed that our results are not so sensitive to price variation.

We solve here the optimization problem for IPM and SRM routers, i.e. maximizing $G$ with arguments of router configurations $m$ and $w$ under the constraint of (10) (the analytical forms of the solution are presented in Appendix C). Though the capacity of computing hardware and the number of routers are expected to have discrete values, we relax this restriction in solving the problem. Figure 10 shows the optimum router configurations, $m^*_I$, $w^*_I$, and $w^*_D$ (superscript $*$ means optimum). In IPM, large memory is required at small data rates, while switching bandwidth is more significant at large data rates\(^1\). Unlike IPM, SRM has no tradeoff between memory and switching bandwidth. The optimum switching bandwidth, $w^*_D$, is constant, because it is determined independent of data rate, as shown in Appendix C. The configurations for IPM and SRM routers provide the optimum router price $c^*$ and the optimum router numbers $N^*$. Figure 11 shows the ratio of optimum router numbers, $N^*_D/N^*_I$.

Figure 10 and Fig. 11 suggest the following implications.

- IPM networks should consist of many narrow-band routers for small data rate applications.
- The small bandwidth implies that router memory does not have to be fast, and we can implement forwarding tables on large and slow memory. We discuss this issue in the next subsection.

\(^1\)Though IPM routers should have the lower limits of the configurations since no commercial product has such small bandwidth as at small data rates or such small memory as at large data rates, it is negligible because the price of either the small bandwidth component or the small memory component is only a small fraction of the optimum router price.
In order to decide which multicast scheme has greater scalability, we take the ratio of $G^*$’s,

$$G^*_D / G^*_I = \frac{c_W|w=w^*_I| + c_M|m=m^*_I| + c_0}{c_W|w=w^*_D| + c_0} \frac{m^*_I}{m^*_D} \frac{l_D}{l_I} \frac{r(h_D + p)}{r(h_I + p)} \frac{\lambda_I}{\lambda_D}. \quad (11)$$

We plot $G^*_D / G^*_I$ of (11) in Fig. 12 (baseline). Since $C$ and $N$ do not appear in (11), we do not need to assign figures to them. At small data rates (publish/subscribe service), an SRM network handles more groups than the IPM equivalent, because IPM routers are very expensive due to their large memory. At large data rates like video streaming, IPM offers slightly higher group numbers than SRM, since the traffic increase incurred by SRM reduces its group number superiority.

4.3 Key Factors on Scalability

The ratio in (11) gives us a way to decide which multicast scheme has greater scalability, but it is hard to identify the key factors that affect the scalability because of the complex form created for the optimization process. In this subsection, we derive a rough approximate form of (11) to identify the key factors and understand their impacts. We focus on the low data rate region in deriving the approximate form.

We first derive approximate forms of $G^*$’s for IPM and SRM.

• IPM. In the low data rate region, $m^*_I$ is approximately constant against $r_p$, and $g^*_I = m^*_I/s$. Since the memory price accounts for a large part of the router price in this region, the number of routers is, roughly, inversely proportional to the memory price and we have $N^*_I \approx C/c^*_M$. From (9), we obtain the following approximation,

$$G^*_I \approx C m^*_I \frac{1}{c^*_M \lambda_I}. \quad (12)$$

We use this approximate form of (12) to estimate the impacts of various factors, and validate the estimation by plots of (11). We begin by understanding the impacts of application factors.

• Data rate. Since (12) is inversely proportional to data rate $r_p$, SRM has greater scalability at lower data rates, as shown in Fig. 12.

• Group size. In (12), $l^*_I$ and $\lambda$ are affected by group size $n$. Since $l^*_I \approx l < k \lambda$ by (6) and (2), (12) is degraded by a factor of $k$ at most when group size $n$ is large. We plot (11) by assuming 10 times larger groups, $n = 1e+3$, in Fig. 12 (large groups). The plot shows that group size

Fig. 11 The ratio of optimum router numbers between IPM and SRM (log-log plot).
has little impact.

We next discuss technology trends and issues that are likely to affect scalability.

- **Moore's law.** The performance and price of computing hardware are, over the long term, improving at exponential rates. If memory and switching components follow parallel trends as before [27], [30], ratios of \( w^*/m' \) and \( c_{W'}/c_{W} \) are unchanged in (12). Therefore, SRM will keep its competitive edge at lower data rates in the future (the plot is omitted due to space limits).

- **Large and cheap memory.** There are some scenarios that change the parallel improvements. Reference [31] argues that reduced latency DRAM would allow forwarding table size to be increased by a factor of ten or more. In addition, IPM routers may use slow and large memory for small data rate services, as discussed in the previous subsection. We replot (11) by assuming 100 times larger memory at the same price, see in Fig. 12 (large memory). The figure shows that SRM may lose its advantage at low data rates (the gain decreases to 1%), since the large memory greatly increases \( m' \) in (12).

- **Delivery tree cache.** We discuss the impact of the delivery tree cache, which was described in Sect. 2.2. In SRM, root routers may be forced to install fast cache memory to store calculated delivery trees. This cache requires the capacity of \( h_1 + d'bl_{D}' \) for each group. We assume that 10% of groups require their trees be cached and a part of the router’s budget is used to buy the cache memory (cache memory size is chosen to minimize \( G_D' \)). We then recalculate \( G_D' \) and plot the results in Fig. 12 (tree cache). The figure shows a significant deterioration in SRM performance. The approximate form (12) does not explain this deterioration directly.

- **Expensive SRM routers.** SRM requires new functionalities, which could make routers more expensive or make \( c_{W} \) larger in (12) due to development cost. We plot (11) by assuming a five fold increase in router price, see Fig. 12 (expensive SRM routers). Even at this unlike router price, SRM does not lose its competitiveness; “five times” is significant in the market.

We conclude that larger memory and tree cache are serious risk factors that could offset the advantage of SRM, though memory and switching components have followed parallel trends historically.

5. Related Work

Several new multicast techniques based on IPM have been proposed. References [9], [10] discussed forwarding state aggregation in an analogy with unicast. Reference [11] compresses forwarding tables by using Bloom filters. In Dr. Multicast [12], small groups are handled as multiple unicasts instead of IPM. Multicast with adaptive dual-state (MAD) [2] introduces scalable overlay multicast for inactive groups, while only a few active groups are maintained by normal IPM. These techniques can reduce forwarding tables by 1/10 at best, which means \( s \) in (11) decreases to 1/10.

Reference [13] is the first research work on SRM to the best of our knowledge. FRM [14] kindled an interest in SRM by proposing a novel encoding scheme that utilizes Bloom filters; it provides great space efficiency and fast operations. In FRM, root routers encode all links of a delivery tree into a Bloom filter. Each link occupies only around 10 bits in the filter, regardless of the original length of the link identifier. Intermediate routers are allowed to check the filter in constant time. FRM also exhibited an advantage in the routing process. An extension of existing unicast routing protocols like BGP and OSPF, FRM avoids the complexity of distributed multicast route computation. In addition, the membership compression technique introduced in FRM leads to the conclusion that its amount of routing information is comparable to that of BGP. Line speed publish/subscribe inter-networking (LIPSIN) [7] introduced special link identifiers that represent a set of frequently used links. This trick can drastically decrease the number of link identifiers to be encoded, though routers are required to maintain the link sets in memory. Reference [32] proposed a fast packet forwarding technique for SRM with Bloom filters. BloomCast [4] proposed a stateless solution that prevents forwarding loops caused by false positives of the Bloom filters.

There are few studies that focus on the scalability of SRM. The above SRM papers ran several computer simulations for evaluation, but they presented no analytical results on scalability. GXcast [17] discussed efficient subgrouping for Xcast, but they focused on forwarding delay since Xcast requires a complicated packet forwarding mechanism unlike SRM.

6. Conclusion

In this paper, we compared the group number scalability of IPM and SRM. We first analyzed the multicast traffic of a single group. SRM is surprisingly inefficient and causes heavy traffic of \( O(\tilde{D}) \sim O(n^{1.6}) \), which is even worse than the unicast traffic, \( O(n) \). The introduction of subgrouping, however, drastically reduces the traffic to \( O(l) \sim O(n^{0.8}) \), which is as efficient as IPM. We next examined the scalability of multicast networks in terms of group numbers supported under the same budget constraint and current specifications of routers on the market. SRM showed much greater capacity than IPM for small bit-rate streams like those in publish/subscribe service, but its advantage can be degraded because of memory issues, such as innovative IPM routers with large memories or the necessity of introducing delivery tree caches to SRM routers. The impacts of these factors are well captured by our approximation.

We hope that SRM will be studied in more detail given the anticipation of a huge number of groups.
Appendix A: Proof of $l_D \approx l$

We assume $n \gg 1$, since subgrouping is used for large groups that should be divided to decrease the traffic. We also assume $d \ll n$, because we do not need to consider inefficient unicast-like transmission such as $d = n$.

Reference [20] derived the following approximate form of (1) for $n \gg 1$,

$$l \approx \sum_{i=1}^{R} \frac{nk_i}{n+k_i}.$$  \hspace{1cm} (A-1)

We next derive an approximate form of $l_D$. We have $[k'/d] \approx k'/d$ for large $i$. In (6), the terms of large $i$ provide a significant contribution. Therefore, the ceilings can be removed from (6). By using (A-1), we have the following approximate form of $l_D$,

$$l_D \approx d \sum_{i=1}^{R} \frac{k_i}{d} \left( 1 - \left( 1 - \frac{1}{k_i/d} \right)^{n/d} \right) \approx \sum_{i=1}^{R} \frac{nk_i}{n+k_i}.$$  

Finally, we find $l_D \approx l$ for $n \gg 1$ and $d \ll n$.

Appendix B: Behavior of $h^*_D$ against $n$

We discuss the header size at the optimum number of subgroups, $d^*$, by using Fig. 4 and Fig. A-1. We assume that the delivery trees depicted in these figures are constructed on the same network (same $k$ and $R$) but the group in Fig. A-1

![Fig. A-1](Image)

An example of a delivery tree that has a group twice the size of that in Fig. 4.

References


is two times larger than that in Fig. 4. Since the half tree in Fig. A-1 is almost identical to the whole tree in Fig. 4, these (half) trees must have a similar number of subgroups of similar size. This implies that packet headers in both figures have nearly the same size. We, finally, make the conjecture that the optimum header size is nearly constant against group size, \( h_D^* = O(1) \).

Figure A-2 plots \( h_D^* \) versus \( n \). The figure demonstrates that \( h_D^* \) remains constant for large \( n \), with some fluctuation due to the ceiling functions in (5).

**Appendix C: Optimum Router Configurations**

As mentioned in Sect. 4.2, the price of computing hardware can be approximated by a cubic function. We define the price functions of memory and switching components as follows,

\[
\begin{align*}
c_M &= m(\alpha_M (m - \beta_M)^2 + \gamma_M), \\
c_W &= w(\alpha_W (w - \beta_W)^2 + \gamma_W),
\end{align*}
\]

where \( \alpha_M, \beta_M, \) and \( \gamma_M \) are determined to fit the market price as in Fig. 8 while \( \alpha_W, \beta_W, \) and \( \gamma_W \) are determined to fit the market price as in Fig. 9.

By solving the optimization problem defined in Sect. 4.2, we obtain the following configurations,

\[
\begin{align*}
m^*_1 &= \left( T^{1/3} + T^{-1/3}\left( \frac{t_0}{3} \right)^2 + t_0/3 \right) \lambda s, \\
w^*_1 &= \left( T^{1/3} + T^{-1/3}\left( \frac{t_0}{3} \right)^2 + t_0/3 \right) lq, \\
w^*_D &= W^{1/3} + W^{-1/3}\left( \frac{\beta_W}{3} \right)^2 + \frac{\beta_W}{3},
\end{align*}
\]

where \( q, t_0, T, \) and \( W \) are given by,

\[
\begin{align*}
q &= r(h_1 + p), \\
t_0 &= \frac{\alpha_M(\lambda s)^2 \beta_M + \alpha_M(lq)^2 \beta_W}{\alpha_M(\lambda s)^3 + \alpha_M(lq)^3}, \\
T &= \sqrt{\left( \frac{t_0}{3} \right)^3 + \frac{T^2}{4} + \left( \frac{t_0}{3} \right)^3 + \frac{T}{2}}, \\
W &= \sqrt{\left( \frac{\beta_W}{3} \right)^3 + \frac{\omega^2}{4} + \left( \frac{\beta_W}{3} \right)^3 + \frac{\omega}{2}}.
\end{align*}
\]

Yohei Katayama received the B.E. and M.E. degrees from Keio University, Kanagawa, Japan, in 2007 and 2009, respectively. Since joining NTT Network Innovation Laboratories in 2009, he has been engaged in R&D on multicast technology and network virtualization technology. His research interests are in architecture design of future networks. He received the research awards of the IEICE Internet Architecture Group in 2008.

Takeru Inoue received the B.E., M.E., and Ph.D. degrees from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2006, respectively. He joined NTT Laboratories in 2000. He is also a researcher at Japan Science and Technology agency. His research interest includes design and control of network systems. He received the best paper award from Asia-Pacific Conference on Communications in 2005. He also received the research awards of the IEICE Information Network Group in 2001, 2004, and 2011. He is a member of IEEE.

Noriyuki Takahashi received the B.E. and M.E. degrees in information engineering from Kyoto University, in 1990 and 1992, respectively. He joined NTT laboratories in 1992. He has been in the NTT Network Innovation Labs. since 2001. His current interests include routing mechanisms, adaptive networking systems, and architecture design of future networks. He is a member of IPSJ, ACM, and IEEE.

Ryutaro Kawamura received B.S. and M.S. degrees in precision engineering and a Ph.D. in electronics and information engineering from Hokkaido University, Japan, in 1987, 1989, and 1996, respectively. In 1989 he joined Nippon Telegraph and Telephone Corporation’ s (NTT) Transport Systems Laboratories. From 1998–1999, he was a visiting researcher in Columbia University. He is engaged in research on network reliability techniques, network control and management, high-speed computer networks, active networks, network middleware and future internet. Currently, he is a director and vice president in NTT Network Innovation Labs. From 2003 to 2013, he is a Board of Director of OSGi Alliance, and from 2005 to 2013 he is a Vice President Asia Pacific.