CSE Articles

Department of Computer Science & Engineering

August 2006

A Framework for Realistic and Systematic Multicast Performance Evaluation

Li Lao
Computer Science Department, University of California, Los Angeles

Jun-Hong Cui
Computer Science & Engineering Department, University of Connecticut, Storrs, jcui@engr.uconn.edu

Mario Gerla
Computer Science Department, University of California, Los Angeles

Follow this and additional works at: https://opencommons.uconn.edu/cse_articles

A Framework for Realistic and Systematic Multicast Performance Evaluation [*]

Li Lao [a] , Jun-Hong Cui [b] , Mario Gerla [a]

[a] *Computer Science Department, University of California, Los Angeles*
[b] *Computer Science & Engineering Department, University of Connecticut, Storrs*

Abstract

Previous multicast research often makes commonly accepted but unverified assumptions on network topologies and group member distribution in simulation studies. In this paper, we propose a framework to systematically evaluate multicast performance for different protocols. We identify a series of metrics, and carry out extensive simulation studies on these metrics with different topological models and group member distributions for three case studies. Our simulation results indicate that realistic topology and group membership models are crucial to accurate multicast performance evaluation. These results can provide guidance for multicast researchers to perform realistic simulations, and facilitate the design and development of multicast protocols.

Key words: multicast, performance evaluation, topology models, group membership models, power law

1 Introduction

Multicasting has been and continues to be a very active research, development, and testing area ever since its inception. In the last two decades, from IP multicast to overlay multicast, numerous multicast schemes and protocols have been proposed [36,7,37,21,22,24,16,28,31,18,29,38,8]. The performance of these approaches and protocols may be significantly influenced by factors such as network topologies and group membership distributions. However, due to a

lack of real data and an absence of a systematic simulation methodology, multicast research usually relies on a variety of commonly accepted but not verified assumptions and models regarding important issues such as group size, membership distribution, and network topology. As a result, simulation results are not comparable and their conclusions may be of limited use.

Similar to other networking areas, multicast research can greatly benefit from realistic models and a systematic evaluation methodology. First of all, the characteristics of network topologies, such as the number of nodes, average node degree, flat or hierarchical structures, and edge distributions, are among the factors that affect multicast performance. For example, since data packets are only replicated at branching nodes, the efficiency of multicast (over unicast) would improve if there are fewer number of non-branching nodes on the paths between multicast group members. Therefore, realistic network topologies are crucial for developing and validating multicasting related protocols and algorithms. In contrast, current multicast researchers often select synthetic topologies arbitrarily for performance evaluation. Second, a realistic model of group membership distribution can have significant impact on the design and development of multicast protocols. It can help to address the scalability issues, and can provide information for reliable multicast protocols to fine-tune their performance. However, a lot of multicast research assumes random group membership or makes other assumptions without verification. Thus, it is very necessary and important to develop realistic models and systematic framework to facilitate multicast protocol design and performance evaluation.

Some prior research works show that unrealistic assumptions could seriously influence the multicast simulation results. For instance, Chuang and Sirbu proposed a power-law relationship to express multicast efficiency in terms of the number of group members [19]. Chalmers et. al. developed a similar metric to measure the gain of multicast over unicast and investigated how receiver distribution and network topologies may affect this metric [14]. An analysis on multicast forwarding state scalability show that forwarding state distribution and state reduction are affected greatly by factors such as network topological properties, group density and membership distribution [48]. It has also been shown by Fei et. al. [27] and Thaler et. al. [44] that multicast state aggregatability heavily depends on group membership models, i.e., how are the members distributed in the network.

As a consequence, some interesting questions that a multicast researcher should ask include: Which (topology and membership) models should I use in my simulations? Do the simulations realistically reflect the performance of a protocol? How could I fairly compare various protocols? To answer these questions, it is crucial to develop a multicast performance evaluation framework so that protocols can be evaluated and compared in a systematic and extensive manner,

and thus facilitate the design of multicast protocols. In this paper, we propose a systematic multicast performance evaluation framework, addressing the above issues by focusing on the "realism" of topology and group membership models. We examine the effect of these models on various multicast performance metrics for three case studies and quantify the "error" due to unrealistic assumptions. The goal of our work is two-fold: on the one hand, we aim to provide guidance for multicast researchers to carry out realistic simulations for the purpose of designing, evaluating and fine-tuning their protocols; on the other hand, we attempt to "validate" the realism of graph (including topology and membership) models from the point of view of multicast performance. From the simulation results based on our proposed framework, we found that unrealistic models do impact the performance evaluation. We identified several topology models that are realistic in terms of multicast performance, and quantified the effects of group member distribution on several performance metrics.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of existing topology generators and group member models. In Section 3, we define our framework for evaluating multicast protocols. In Section 4, we present simulation results with different topologies and group member distributions for three case studies. Finally, we summarize our contributions in Section 5.

2 Related Work

In this section, we discuss related work in areas of topology generators, group member models, and multicast performance evaluation. These results have greatly influenced our work.

2.1 Topology Generators

How to correctly model and generate network topology is a fundamental research problem. The design and development of new network protocols involve extensive testing on the feasibility of these protocols. However, due to the difficulty of performing experiments on large-scale Internet, researchers often turn to analysis and simulations to predict or estimate the protocol performance. As a result, a number of topology models and generators have been proposed to produce synthetic graphs for the purpose of simulations. Network topologies that resemble the structure and properties of the Internet would accurately guide the design and testing of various network protocols, whereas unrealistic topology models may lead researchers to draw incorrect conclusions.

Early work such as Waxman Model [46] relies on random graphs to generate network topologies. In this model, the nodes in the network are distributed randomly on a two-dimensional grid. An edge between two nodes is added to the graph according to a probability function of the Euclidean distance between them. While Waxman model produces a flat network, GT-ITM package [13] and Tiers [3] attempt to capture the hierarchical structure of the Internet. By using five random graph models (including Waxman model), GT-ITM Transit-Stub Model produces two-level graphs representing transit domains and stub domains. In Tiers, three levels of hierarchy, namely, WAN, MAN and LAN, are created. All the nodes in a single domain are joined by a minimum spanning tree, and redundant intra- and inter-domain edges to closest nodes are added to satisfy node degree.

The seminal work by Faloutsous et. al. [25] showed empirically that certain properties at Autonomous System (AS) Level can be described with power-laws. These properties include node degree, complementary cumulative distribution function of node degree, graph eigenvalues, and the total number of pairs of nodes within a certain number of hops. Following this work, a number of topology generators [9,34,32,12] have been proposed to incorporate these power laws either explicitly or implicitly. Barabasi and Albert provide a model to explain the origin of the degree power law. In their model, they incrementally generate nodes and attach new nodes to existing ones with a probability proportional to their degrees (also called linear preference model) [9]. Boston university Representative Internet Topology gEnerator (BRITE) extends Barabasi-Albert model by combining preferential attachment with a heavy-tailed distribution of node placement [34]. Based on the observation that new ASs have a stronger preference to connect to high degree ASs than linear preference model [15], Generalized Linear Preference model (GLP) is proposed with a modified probability function for new node attachment [12]. Internet Topology Generator (Inet) makes use of both explicit and implicit approaches: it uses the power laws explicitly to generate outdegrees for all nodes, builds a spanning tree among nodes with degree higher than 1, attaches remaining nodes to the spanning tree, and connects nodes with unfilled degrees according to the linear preference model [32].

2.2 Group Member Models

Previous research work on group member models addresses the temporal and spatial properties of group members. [6,30] model join and leave behaviors in single-source applications over MBone and in multiple-source networked games, respectively. In both works, the authors found that time-of-day effect varies the number of participating members. In addition, the inter-arrival times fit well to an exponential distribution, and membership duration follows

exponential distribution for shorter sessions and Zipf distribution for longer sessions.

In terms of spatial properties (i.e., group membership), Phillips et. al. proposed an Affinity/Disaffinity model to control the likelihood that group members are clustered or spread out [39]. However, this model is not able to evaluate the affinity/disaffinity of real groups or groups generated with other models. Lucas et. al. proposed several distance- and expansion-based metrics to better characterize group clustering, but they did not provide algorithms to generate group membership according to these metrics. Recently, a comprehensive and systematic group membership model GEneralized Membership Model (GEM) is developed by Cui et. al. [20]. GEM models group spatial properties with respect to member clustering, group participation probability, and pairwise correlation in group participation. The authors show through measurement that real group communication applications such as audio/video conferences on MBone and network games exhibit distinct behaviors in terms of spatial group properties. Their group generator is able to create groups preserving those properties.

In this paper, we mainly focus on the impact of group membership on multicast evaluation.

2.3 Multicast Performance Evaluation

There has been some research work on evaluating multicast performance with respect to end-to-end delay, tree cost, traffic concentration, multicast efficiency, and state aggregatability.

In [47,11], the performance trade-offs between source-based trees and core-based trees are investigated regarding end-to-end delay, tree cost, control overhead, traffic concentration, and routing table size, for uniform randomly distributed members. In [47], simulations are carried out over an early ARPAnet with 47 nodes and Waxman graphs, whereas in [11], a hierarchical but unrealistic topology model is adopted.

After Chuang and Sirbu discovered multicast tree cost varies at the 0.8 power of group size [19], Phillips et. al. provided one possible explanation for this power law and found that affinity/disaffinity greatly affects multicast tree cost [39]. Chalmers and Almeroth show that ignoring certain temporal and spatial group properties does affect the efficiency gain of multicast (versus unicast) [14]. A recent work [33] on small-world characteristics of the Internet reveals that IP multicast tree size depends on the small-world behavior. In [23], Fahmy and Kwon characterize application level multicast trees on topology with power-law and small-world properties for uniform and non-uniform host

distribution.

Perhaps the work most closely related to ours is [48,49,40]. Wong et. al. study the multicast forwarding state scalability with a variety of real and synthetic topologies and membership models [48]. In [49,40], the authors focus on characterizing and selecting realistic graph models. They compare the generated graphs using a set of graph metrics and identify the impact of these topologies on multicast protocol design. However, their work concentrates on early graph models such as Waxman, Transit-Stub, and Tiers, with an assumption of uniform random group member distribution.

Our work is distinct from previous work in that we attempt to propose a systematic framework and aim to evaluate multicast performance in a more comprehensive and realistic way: 1) we identify a wide variety of multicast performance metrics; 2) we investigate the impact of network topologies, especially those produced by power-law topology generators, since they resemble the Internet better than previous ones but their impact on multicast has been rarely studied; 3) we evaluate multicast performance with a generic membership model (which contains the commonly used uniform membership model) and examine how different parameters affect various multicast metrics; 4) we also conduct a case study of application layer multicast, on which the impact of network topologies and group membership models has not been explored in the literature. In a nutshell, we want to identify realistic models and parameters with which the multicast protocols exhibit similar behaviors as if they were tested in the Internet for real group communication applications.

3 The Systematic and Realistic Performance Evaluation Framework

In this section, we define our framework for systematic and realistic multicast performance evaluation. The framework consists of three basic elements: Models and Parameters MP, Metrics M, and Protocols PR. In other words, we want to study what is the performance of a protocol PR regarding metrics M for given MP? We present these three elements one by one in the following subsections.

3.1 Models and Parameters

We first revisit existing topology and membership models and their assumptions. Based on the realism of these models, we choose appropriate ones to incorporate into our performance evaluation.

Topology Models and Parameters The Internet consists of many autonomously administered domains or Autonomous Systems (AS). In networking research, two types of network topologies are typically considered: AS-level and router-level. Researchers could obtain real Internet topologies (AS-level or/and router-level) from several Internet measurement and research centers/projects, such as CAIDA [1], NLANR [2], and Route Views [4]. Considering the extremely large size of the real Internet-wide router-level topology (For example, a router-level graph measured by CAIDA includes 192244 nodes, 636643 directed links, and 609066 undirected links) and the fact that most of multicast protocols are designed for single domains, people usually tend to use router-level topologies in a single big domain. Some frequently used are AT&T, MCI, and Internet 2 (Abilene), etc.

In the literature, there are a variety of graph models for generating topologies. They can be categorized as follows: *flat random models* (Waxman), *hierarchical random models* (Transit-Stub and Tiers), and *power-law models* (Barabasi-Albert, BRITE, GLP and INET, to name a few). It has been shown in [43] that power-law models better represent the large-scale structure of measured topologies with respect to topological metrics. On the other hand, Transit-Stub hierarchical random model could generate network topologies more resembling the router-level topologies for single domains.

In our work, we will study real network topologies and generated network topologies at both AS-level and router-level. For AS-level topologies, we use real data from the University of Oregon Route Views Project [4]. These data have been frequently used by researchers and they have instances spanning over a period of 5 years. In these topologies, we treat the Internet as an undirected graph whose nodes are AS routers and whose edges are inter-domain connections with unit weight. The network size of these graphs varies approximately from 3000 to 11000 nodes. We compare topologies generated by some typical power-law models (such as Barabasi-Albert, BRITE, GLP and INET) with real Internet instances. We vary the network size of the synthetic graphs in the same range as real graphs. For router-level topology, we employ AT&T (one of the biggest tier-1 ISPs in the world) backbone network. Using Transit-Stub and Tiers models, we generate topologies resembling the real AT&T topology by tuning the model parameters.

Membership Model and Parameters As mentioned earlier, a common assumption regarding group membership in multicast simulation is that group members are uniformly distributed in the network, i.e., all nodes have same group participation probabilities. In our work, we challenge this assumption, and propose a generic group participation probability model based on the membership measurement results in [20]. Before presenting the model, we first give a formal definition of group participation probability.

Group Participation Probability Given a network $G(V, E)$, for any node i, where $i \in V$, group participation probability p_i is defined as the probability that node i joins a group.

It has been shown in [20] that the distribution of group participation probability for applications such as MBone video conference and net games can be approximated by normal distribution with a probability density function (PDF):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{1}$$

where μ is the mean and σ is the standard deviation. Note that, when σ approaches 0, $f(x)$ becomes a $\delta(x - \mu)$ (Dirac delta function), thus the group participation probability for all the nodes are the same as μ; in other words, the group is uniformly distributed. Therefore, we can say that the above normal distribution model includes the uniform distribution model.

The suggested normal distribution model is powerful, however, due to the valid range of group participation probability (i.e., $0 \leq p_i \leq 1$, $\forall i \in V$), it is hard to directly harness the range of generated probabilities for given μ and σ. To conquer this problem, we propose to use truncated normal distribution [10]. A truncated normal distribution at an interval of $[0, 1]$ has the following probability density function (PDF):

$$f_{tr}(x) = \frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{1-\mu}{\sigma}) - \Phi(\frac{-\mu}{\sigma})}, \tag{2}$$

where $0 \leq x \leq 1$; μ and σ are the mean and standard deviation of the non-truncated normal distribution respectively; $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $-\infty \leq x \leq \infty$; and $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$.

It is easy to derive the mean of the proposed truncated normal distribution:

$$E_{tr}[X] = \mu - \sigma \frac{\phi(\frac{1-\mu}{\sigma}) - \phi(\frac{-\mu}{\sigma})}{\Phi(\frac{1-\mu}{\sigma}) - \Phi(\frac{-\mu}{\sigma})}. \tag{3}$$

Clearly, in most cases, $E_{tr}[X]$ is not equal to μ. Thus, we have to control group size based on $E_{tr}[X]$ instead of μ.

Control the Group Size Once we know the group participation p_i for node i (where $i \in V$), the average group size can be estimated to be $\sum_{i \in V} p_i$. However, sometimes, we want to control group size in our simulations. A suggested method in [20] is to recompute the probabilities under the given condition,
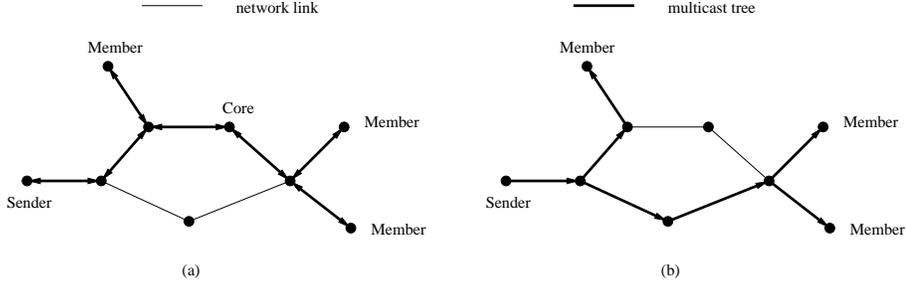
Fig. 1. A comparison of (a) core based tree and (b) source-based tree.

i.e., bounded group size g: p_i is recomputed as $p_i \, [(\sum_{i \in V} p_i)/g]$. In this paper, we adopt this approach to control group size.

Based on the above group membership model, we generate a variety of groups by changing the values of μ and σ, and quantify to what extent these parameters affect the performance of different group communication applications.

3.2   Case Studies

It has been shown in previous work that unrealistic assumptions or models may affect multicast protocol performance in certain aspects. In our work, we consider three case studies, namely, the properties of Source Based Trees (SBT) vs. Core-Based Trees (CBT), multicast state scalability, and application layer multicast protocol comparison (Narada vs. NICE). In our opinions, they are very fundamental problems or protocols in multicast research, and are likely to be sensitive to topologies and/or member distribution. Moreover, the results of these case studies could potentially affect the design of multicast protocols in the future.

SBT vs. CBT There are mainly two flavors of multicast routing algorithms: Source Based Tree and Core Based Tree. The SBT approach creates a shortest path tree rooted at each source, while CBT approach produces a single bi-directional tree which carries the data packets of all sources. Generally speaking, SBT has the advantage of shorter path length, but CBT introduces less multicast forwarding state since it is shared among all the sources of a group. The difference of these two kinds of multicast trees are shown in Fig. 1. Intuitively, the routing performance of SBT and CBT are easily affected by the variation of topologies and group distribution, so we want study their sensitivity to different models.

Multicast State Scalability Due to the lack of hierarchical address allocation and routing mechanisms, multicast faces serious forwarding state scalability problem because each relaying node needs to maintain state information per group or per source/group pair. Currently, the solutions to this problem

can be classified into intra- and inter-group state aggregation. The *intra-group aggregation* approaches rely on aggregating states for individual groups, such as removing state information at non-branching nodes [45,42]. The *inter-group aggregation* schemes exploit the correlation of forwarding state among groups. For instance, aggregated multicast forces multiple groups to share an aggregated multicast tree [26], whereas [41,44] suggest to aggregate multiple multicast addresses with the same prefix in a router into a single address with mask. Considering that the number of branching nodes and multicast tree aggregation are both closely related to network topologies and group distribution, we decide to analyze the intra- and inter-group state aggregatability.

Application Layer Multicast: Narada vs. NICE Recently application layer multicast has emerged as a new multicast architecture in the Internet. It implements multicast-related features exclusively at end systems, and does not require infrastructure support from intermediate nodes (such as routers or proxies). Data packets are transmitted between end hosts via unicast, and are only replicated at end hosts. Narada [18,17] and NICE [8] are two representative application layer multicast protocols. In Narada [18,17], end hosts periodically exchange group membership information and routing information, build a mesh based on end-to-end measurements, and run a distributed distance vector protocol to construct a multicast delivery tree. This protocol mainly targets at applications with small and sparse groups, such as audio and video conferences. NICE [8], on the other hand, is designed to support applications with very large receiver sets and relatively low bandwidth requirements. It recursively arranges group members into a hierarchical overlay topology, which implicitly defines a source-specific tree for data delivery. Some important questions to ask are: to what extent NICE is more scalable than Narada? Is NICE always a winner for all network topologies and group membership models? In this paper, we will apply our evaluation framework and compare the performance of these two protocols.

3.3 Metrics of Interest

We select a set of performance metrics that are frequently used in multicast performance evaluation and that explore diverse aspects of multicast protocols: routing performance, multicast efficiency and state scalability. We also choose some important metrics for application layer multicast. We want to investigate the sensitivity of these metrics to different topology and membership models.

Routing Performance We choose tree cost and end-to-end delay to measure the routing performance of a protocol. *Tree cost* is defined as the total number of links in a multicast distribution tree, and it represents the amount of network resource usage. *End-to-end delay* is measured as the average delay

from any source to any receiver, which is important for real-time interactive applications. Since the topologies we study do not consider link weight, the delay of a path is assumed to be proportional to the path length.

Multicast Efficiency *Multicast efficiency* is normally defined as:

$$eff = \frac{L_m}{L_u}, \tag{4}$$

where L_m is the total number of multicast links in distribution trees, and L_u is the total number of unicast hops. This metric has been widely studied since Chuang-Sirbu power law was proposed, which states that multicast efficiency (in a slightly different representation) varies at the 0.8 power of group size [19]. Previous work show that efficiency and its power law exponent do vary with different group distribution [39,14]. Therefore, we want to examine the values of efficiency as well as its power law exponent.

State Scalability To evaluate the multicast state aggregatability, we define *state reduction ratio SRR* as follows:

$$SRR = 1 - \frac{S_{agg}}{S_{orig}}, \tag{5}$$

where S_{agg} and S_{orig} are the number of state entries with and without aggregation, respectively. In addition, to evaluate the extent of tree aggregation for aggregated multicast, we define *aggregation degree AD* as the ratio of the number of multicast groups N_g to the number of aggregated trees N_t:

$$AD = \frac{N_g}{N_t}. \tag{6}$$

Link Stress In application layer multicast, end hosts use overlay links to deliver data, so the same data packets may traverse an underlying physical link multiple times. *Link stress* is defined as the number of identical packets delivered on a physical link, and it is a measure of the efficiency of an application layer multicast protocol.

Control overhead Unlike IP multicast, which stores the group membership and multicast tree information in routers, application layer multicast requires end hosts to cooperatively maintain group membership and construct multicast trees. Therefore, the communication overhead among group members may grow explosively as group size increases. We use the sum of the number of links each control packet traverses to estimate the *control overhead*.

4 Simulation Results

In this section, we show through simulation studies the effect of different network topologies and group characteristics on multicast performance (namely, multicast routing algorithms, state scalability and application layer multicast).

4.1 Simulation Setup

In our simulations, we use both *AS-level* and *router-level* network topologies to examine the sensitivity of multicast performance. For AS-level topologies, we select a series of AS level Internet graphs with a wide range of graph sizes. Each instance is named using its date, eg, the instance collected on May 07, 2001 is named I010507, where "I" stands for "instance". We use real Internet instances from November 1997 to March 2003 in our experiments. Since AS-level topologies have been shown to follow power laws, we also generate graphs with size comparable to Internet instances using power law graph models Inet, Barabasi-Albert and GLP. Note that Barabasi-Albert (BA) model has two options (random and heavy-tailed distributions) for node placement, we test the performance of both options. On the other hand, for router-level topologies, we use a real network, AT&T IP backbone [5], which consists of 123 nodes in totoal: 9 gateway routers, 9 backbone routers, 9 remote GSR (Gigabit Switch Routers) access routers, and 96 remote access routers. Based on this graph, we generate a series of random graphs with similar graph size and average node degree using structure-based topology generators, such as Transit-Stub and Tiers. As a further comparison, we include Waxman model, a flat random model that captures neither structural properties nor power laws, for both AS-level and router-level graphs. To focus on the effect of network topologies, we generate groups using uniform distribution.

To evaluate the impact of group membership distribution, we use the proposed truncated normal distribution model to generate group members on a small network with 54 routers, which is abstracted from AT&T backbone network aforementioned. In this simplified network, the gateway and backbone routers remain intact, whereas the remote access routers attached to the same gateway or backbone router are contracted into one node. In addition, we create a neighbor node for each gateway router in the backbone, since gateway routers represents connectivity to other peering networks and/or Internet public exchange points. For group membership generation, we vary μ and σ of the non-truncated normal distribution to produce sets of group participation probability p_i for all the nodes in the network, explained as follows. We first fix σ to 0.1, and increase μ from 0.4 to 0.8, which is equivalent to increasing group size. Then we fix μ to 0.5 and vary σ from 0 to 2. In this way, the

average group size is fixed, but the distribution of participation probability changes from uniform to skewed distribution, that is, some nodes tend to join groups with a higher probability than others. Note that we need to control group size to meet the expected value since $E_{tr}[X]$ is not equivalent to μ. For each combination of μ and σ, we generate 10 instances of group participation probabilities $\{p_i\}$ for every node i. Then for each instance, we generate 10000 multicast groups and measure the performance.

4.2 Multicast Routing Algorithms

We first compare SBT and CBT on real and generated graphs with respect to cost ratio and delay ratio, and evaluate their multicast efficiency. Cost (or delay) ratio is defined as the ratio of tree cost (or end-to-end delay) of SBT to that of CBT. For multicast efficiency, CBT and SBT have similar results; therefore, we only show the efficiency of CBT unless otherwise specified. Note that we also compared the control overhead ratio of SBT vs. CBT, which has similar trend as cost ratio and is thus not shown here.

In the CBT approach, we choose as the core one of the member nodes uniformly. In the remainder of this section, we will evaluate these three metrics with regard to AS-level topologies, router-level topologies, and group membership distribution. The reason we use AS-level topologies is two-fold: first, we want to study the topology sensitivity of multicast algorithms, so we intend to use as diversified topologies as possible; second, using AS-level topologies, we can investigate the performance issue of inter-domain multicast routing algorithms.

4.2.1 AS-level Topologies

For AS-level topologies, we conduct two series of experiments for each metric. *Series A:* since the group size of a multicast group is an important parameter, in this series, we fix graph size and vary the group size. We use real Internet instance I980124 (of approximately 3200 nodes) and generate topologies with comparable size, and vary group size from 0.125% to 32% of the graph size, or from about 4 to 1000 nodes. For each group size, we repeat the experiment 100 times. *Series B:* in this series, we fix the group size to 16 members, and vary the graph size. We select a series of Internet instances at regular time intervals of increasing size from I980124 to I010507, and produce a set of graphs with the same size from each of the topology generators mentioned above. Then we compare the trends of the metric variation on real and generated graphs. We also test other group size, and the results are consistent.

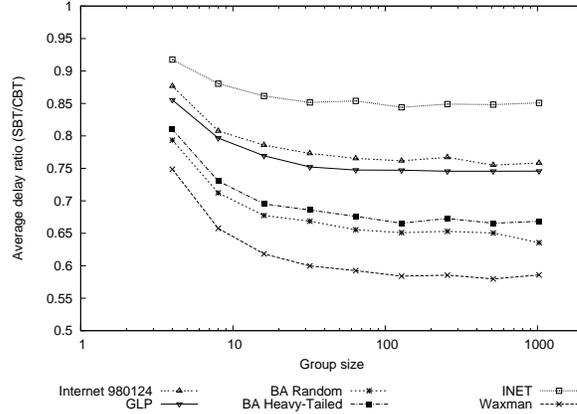**Delay ratio** In Fig. 2, we plot the average end-to-end delay ratio versus the

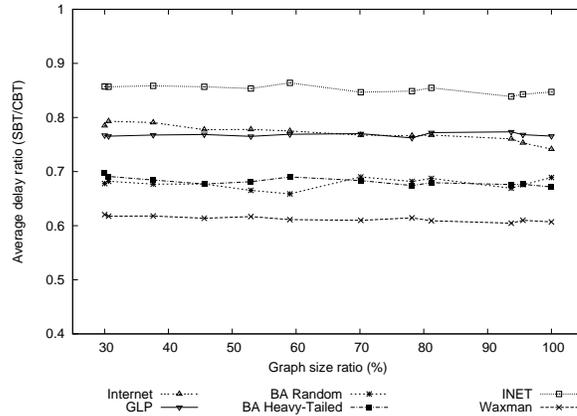Fig. 2. Series A: Delay ratio of SBT/CBT versus group size for AS-level graphs.

Fig. 3. Series B: Delay ratio of SBT/CBT versus graph size for AS-level graphs.

size of the multicast group. We observe that topology does affect delay ratio significantly: the delay ratio of the graphs generated by GLP have average percent deviations within 5% of the delay ratio of Internet instance I980124, whereas the remaining graphs deviate from this real graph for more than 10%. Waxman shows the highest deviation from the Internet graph. The 95% confidence intervals for all points are less than 0.01.

Furthermore, we plot the average end-to-end delay ratio with increased graph size in Fig. 3. For simplicity, we use the ratio of graph size to the size of the largest Internet instance we collected (I010507 with about 11000 nodes) as X-axis in the figure. This graph leads to the same conclusions as Fig. 2: the delay ratio in different topologies of the same size may vary greatly (e.g., from 0.60 to 0.87). It is also surprising to observe that, for each topology generator, the difference in delay ratio between generated graphs and the Internet graphs remains rather consistent, irrespective of the graph size.

Cost ratio In contrast, cost ratio exhibits very different behaviors from delay ratio in terms of topology sensitivity (Fig. 4). For all graphs, cost ratio
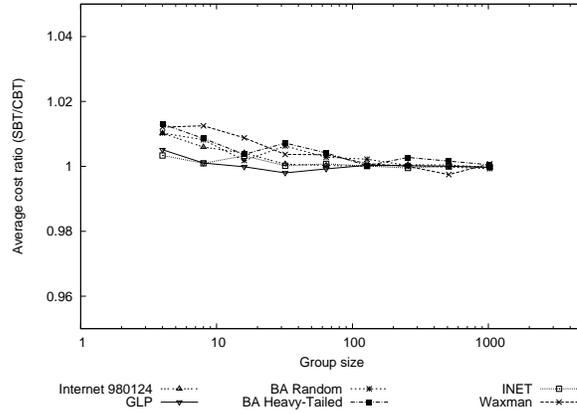
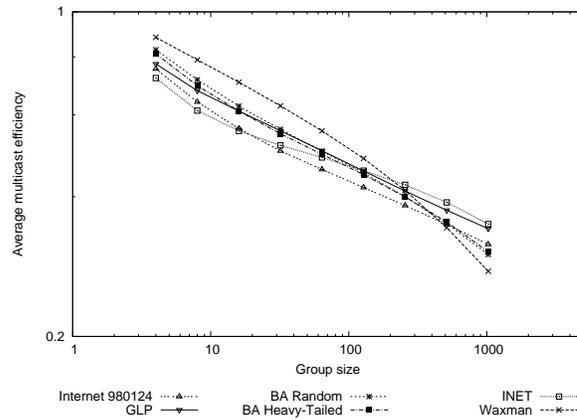Fig. 4. Series A: Cost ratio of SBT/CBT versus group size for AS-level graphs.

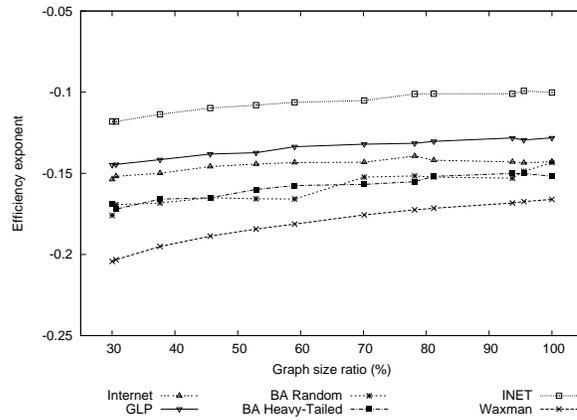Fig. 5. Series A: Multicast efficiency of CBT versus group size for AS-level graphs.

Fig. 6. Series B: Multicast efficiency of CBT versus graph size for AS-level graphs.

fluctuates between 0.99 and 1.02, and most data points of generated graphs fall within the 95% confidence limits of the real graph. We observe the same trend for cost ratio as graph size varies, which is not shown here due to space limitation. Hence, we conclude that cost ratio is not so sensitive to topologies as delay ratio.

Table 1
Series A: Multicast efficiency comparison for AS-level graphs.

Method	Exponent	Corr. Coeff.
Internet I980124	-0.150	0.995
GLP	-0.145	0.999
BA Random	-0.176	0.999
BA Heavy-Tailed	-0.169	0.998
Inet	-0.118	0.991
Waxman	-0.204	0.993

Efficiency As shown in Fig. 5, the efficiency metric decreases with group size, indicating that multicast becomes more efficient than unicast. This is reasonable: as group size increases, more duplicate data packets are transmitted on each link for unicast, thus multicast becomes more efficient compared with unicast. Note that a smaller value of the efficiency metric is equivalent to improved multicast efficiency. For all topologies, multicast efficiency conforms to Chuang-Sirbu power law, even though Waxman model shows a trend different from the remaining graphs. However, the absolute values of the efficiency metric for the synthetic topologies vary within $\pm 20\%$ of the Internet topologies. As a further comparison, we present the efficiency exponents and correlation coefficients of this power law for different graphs in Table 1. The table suggests that both multicast efficiency and its power-law exponent are sensitive to the topology characteristics. For instance, GLP matches the real instance I980124 very well, and Inet and BA topologies also follow I980124 closely with an average percent deviations of approximately $6 - 8\%$. On the other hand, the efficiency exponent of Waxman is not close to that of Internet at all.

Mieghem et. al. find out that the efficiency exponent is not a constant, but slowly increases with the graph size [35]. We calculate the efficiency exponents for graphs with different sizes and plot the results in Fig. 6. All these topologies have a steady increase in efficiency exponent as the graphs are further enlarged. In addition, this figure is consistent with earlier observations that GLP has the closest multicast performance as the real Internet instances, and Waxman random graphs show dramatically different behaviors from Internet topologies.

4.2.2 Router-level Topologies

In router-level topologies, the graph size is fixed at approximately 120, and we compare the performance of CBT and SBT as group size increases. Similar to AS-level topologies, cost ratio is not sensitive to different topologies, so we omit it from further discussion.
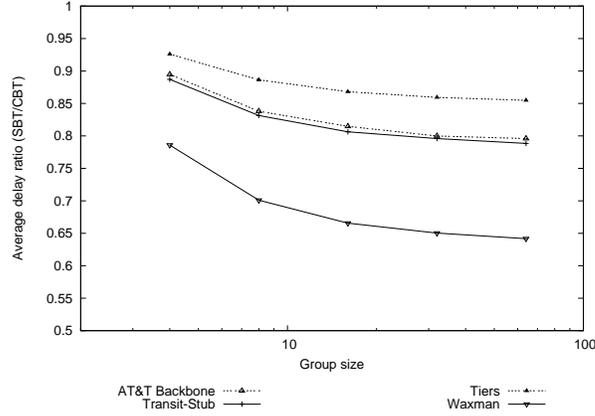
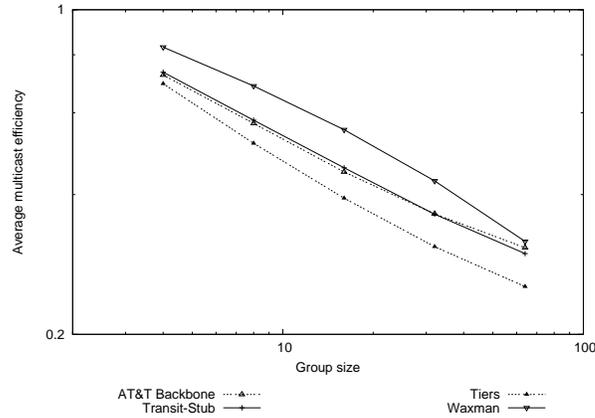Fig. 7. Delay ratio of SBT/CBT versus group size for router-level graphs.

Fig. 8. Multicast efficiency of CBT versus group size for router-level graphs.

Delay Ratio As shown in Fig. 7, delay ratio decreases with increased group size, and its value varies among different topologies. The average percent deviations of Transit-Stub, Tiers and Waxman graphs from AT&T topology are approximately 0.8%, 6.1%, and 17.5%, respectively.

Efficiency We also plot multicast efficiency vs. group size in Fig. 8, and observe similar trend in the router-level topologies as in AS-level topologies: regardless of the topologies, the efficiency metric decreases as group size grows; however, there is some variation in the metric when topologies are changed. Again, among the synthetic graphs, Transit-Stub produces the best match to the real graph.

4.2.3 Group Membership Distributions

We measure the same metrics for groups generated from truncated normal distribution as explained in Section 4.1. The simulation results indicate that these metrics are not significantly affected by group membership distributions for the parameters under investigation, except for multicast efficiency, which is
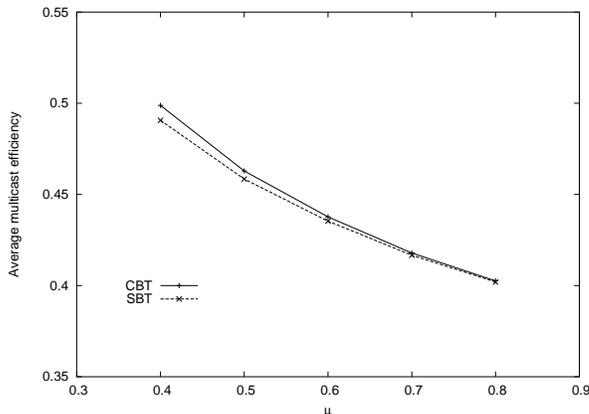
Fig. 9. Average multicast efficiency for multicast groups with $\sigma = 0.1$.

sensitive to μ when σ is fixed. The reason behind this is simple: μ determines expected group size. When we raise the value of μ, group size grows on average; consequently, multicast efficiency improves, which translates into lower values for the efficiency metric (as illustrated in Fig. 9).

4.3 Multicast State Scalability

Recall that there are two general approaches to improve multicast state scalability: intra-group aggregation by removing multicast forwarding state at non-branch nodes, and inter-group aggregation by forcing multiple multicast groups with similar group members to share one aggregated tree. Since multicast state scalability problem is exacerbated in backbone domain, where hundreds or even thousands groups may potentially go through, we focus on small topologies (e.g., the abstracted AT&T network as explained in Section 4.1) to represent such backbone networks. Thus, in this section, we mainly investigate to what extent group membership may impact multicast state aggregation. We assume SBT is used for inter-group aggregation. We evaluate SRR for both approaches and AD for inter-group aggregation, under two cases illustrated in Section 4.1: 1) when μ is fixed and σ is varied; or 2) when σ is fixed and μ is varied.

4.3.1 Intra-group State Aggregation

The SRR of intra-group state aggregation is proportional to the percentage of branching nodes on the tree. In the first case, when μ is fixed, SRR remains relatively stable, since the average group size is also fixed. In the second case, as μ is increased, there are more members in each group on average. When a node joins a group, it needs to be grafted onto an on-tree node through the shortest path to the source or to the core node, which may increase the number
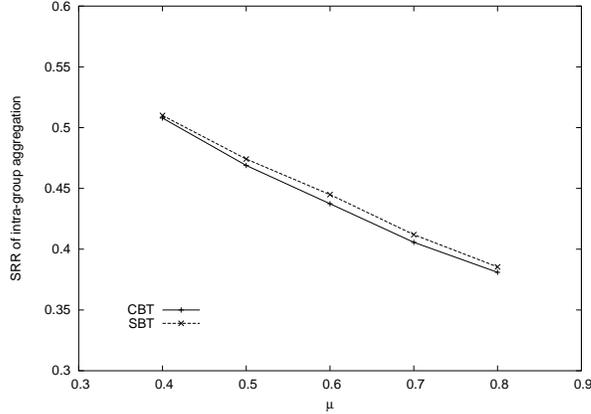
18

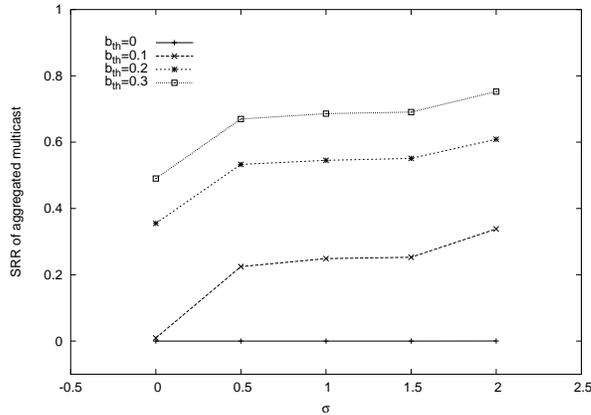Fig. 10. SRR of intra-group state aggregation for multicast groups with $\sigma = 0.1$.

Fig. 11. SRR of aggregated multicast for multicast groups with $\mu = 0.5$.

of branching nodes. As a result, the fraction of branching nodes increase in the network, yielding less state reduction. As demonstrated in Fig. 10, when μ increases from 0.4 to 0.8, SRR reduces from 0.51 to 0.39.

4.3.2   Inter-group State Aggregation

Inter-group aggregation (e.g., aggregated multicast) aggregates multicast groups with similar sets of members into one tree. Generally, there are two types of matching between a group and a multicast tree. If every tree leaf is a group member, then the tree is a "perfect match" for the group; otherwise, the tree is a "leaky match" for the group, since some multicast data will be delivered to non-member leaves. Obviously, there is a trade-off of aggregation vs. bandwidth consumption: leaky match allows more groups to share an aggregated tree, but more bandwidth is consumed. For this reason, a bandwidth waste threshold b_{th} is adopted in the group-tree matching algorithm: the higher b_{th} is, the more bandwidth waste is allowed and the fewer trees are needed to cover all groups. Thus, we use b_{th} as an additional parameter in our simulations.
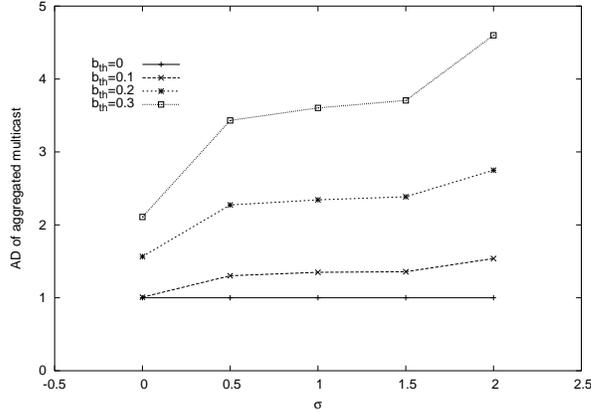
19

Fig. 12. AD of aggregated multicast for multicast groups with $\mu = 0.5$.
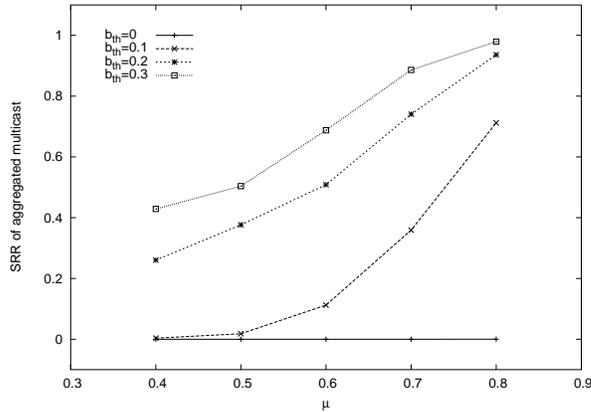

Fig. 13. SRR of aggregated multicast for multicast groups with $\sigma = 0.1$.

For inter-group aggregation, if some combinations of group members occur very often, the number of aggregated trees can be significantly reduced. Thus, we expect its performance to be strongly affected by group member distribution across multiple groups.

In the simulations, we first set $\mu = 0.5$ and start with uniform group member distribution (i.e., $\sigma = 0$). As σ increases, the group participation probabilities become more skewed: some nodes have higher probability to join groups, and others have lower probability. Therefore, some group member combinations have higher probability to occur, which means more multicast groups can be aggregated onto each tree. Fig. 11 and 12 confirm this intuition by showing that SRR and AD both increase in these two topologies, especially for higher bandwidth waste threshold b_{th}, since bandwidth is traded to allow more groups to be aggregated onto the same tree.

As shown in Fig. 13 and 14, when $\sigma = 0.1$, SRR and AD improve with μ. Intuitively, when μ is very large, almost all nodes will join every group, so large groups will occur very often and a small number of large trees suffice for
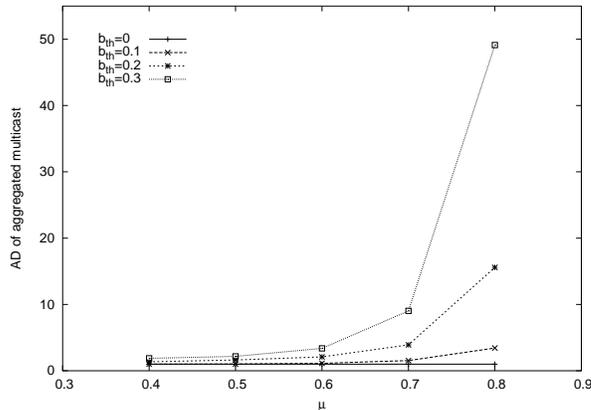
20

Fig. 14. AD of aggregated multicast for multicast groups with $\sigma = 0.1$.

most of these groups. In contrast, for relatively small μ, the nodes have smaller participation probabilities, so it is less likely that some particular combinations of group members will appear much more often than the rest. As a result, SRR and AD are much higher in the former case.

4.4 Application Layer Multicast

In this section, we use a packet level simulator to evaluate two representative application layer multicast protocols, NICE and Narada, with respect to multicast tree cost, end-to-end delay, link stress and control overhead. For each metric, we use the ratio of NICE vs. Narada to compare their performance. It is worth noting that our goal is not to compare which protocol is better; instead, we want to investigate whether the performance of application layer multicast protocols is affected by topology models and group membership models. We test these two protocols on both AS-level and Router-level topologies with uniform group distribution. At first glance, it may seem unreasonable to use AS topologies for application layer multicast performance evaluation. Nevertheless, we make a similar argument as in Section 4.2: we want to study their performance in diverse environments and explore their potential of being used in a hierarchical fashion to connect cluster heads residing in each AS. We also vary the distribution of the group participation probability to compare these two protocols. However, in spite of various group member distributions, simulation results indicate that the above performance metrics are rather stable for both protocols. Accordingly, we only show results for uniformly distributed group members in this subsection.
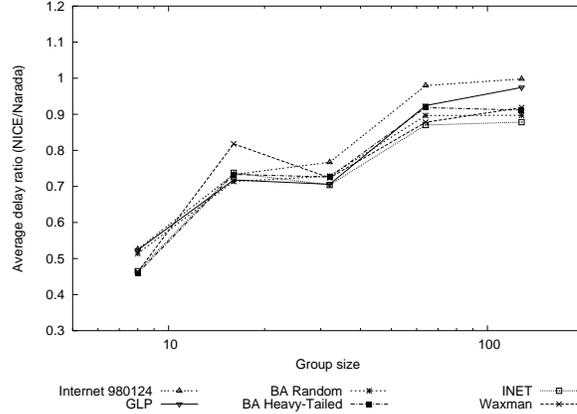
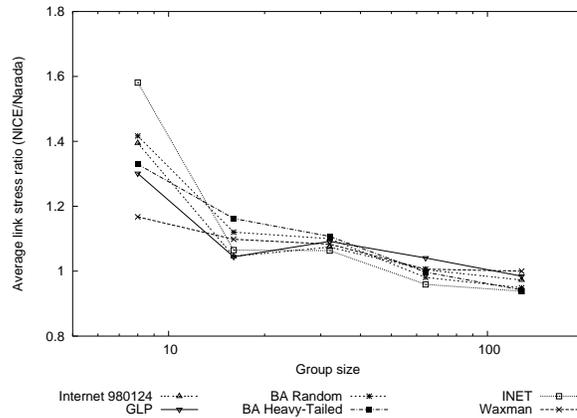Fig. 15. Delay ratio of NICE/Narada for AS-level graphs.

Fig. 16. Link stress ratio of NICE/Narada for AS-level graphs.

4.4.1 AS-level topologies

For consistency, we select the Internet instance I980124 and synthetic graphs with similar size in the simulations. We intend to focus on the inter-domain connections, so the group members are randomly selected from each AS router. The group size is varied from 4 to 128, which can be accommodated by both protocols. A member is uniformly randomly assigned as source. Since Narada is designed for small groups with tens of members like video-conferences and NICE is designed for large groups, we expect NICE to outperform Narada when group size is relatively high. This is indeed the case for most of the metrics we evaluated.

Delay Ratio As depicted in Fig. 15, the delay ratio of NICE vs. Narada differs in various topologies, with the highest deviation from real Internet around 12%. Among the topology models, GLP and BA perform the closest to the real graph, and Waxman deviates the most from the Internet topology.

Link Stress Ratio Fig. 16 demonstrates that topology models have some impact on link stress ratio, especially when group size is small. For group size
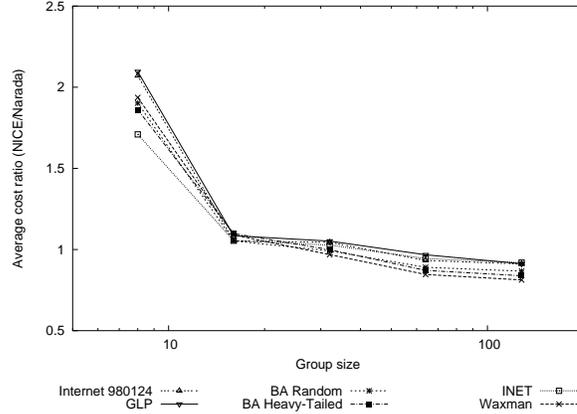
22

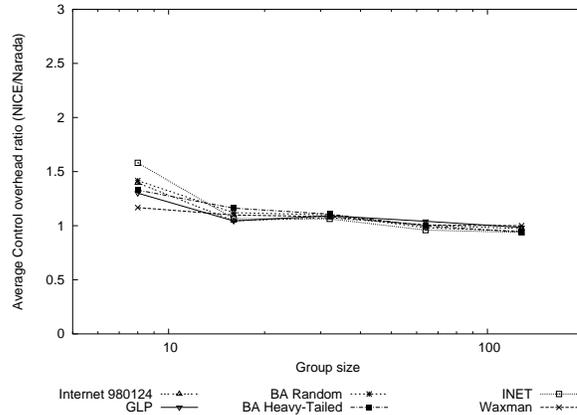Fig. 17. Cost ratio of NICE/Narada for AS-level graphs.

Fig. 18. Control overhead ratio of NICE/Narada for AS-level graphs.

of 4, the average link stress ratio on synthetic graphs ranges from 1.30 to 1.58, while that of the Internet is approximately 1.40.

Cost Ratio and Control Overhead Ratio Finally, from Fig. 17 and 18, we found that the cost ratio and control overhead ratio are not very sensitive to AS-level topology models. These two metrics converge for slightly large groups.

4.4.2 Router-level topologies

We use similar router-level topologies as in Section 4.2, AT&T backbone, Transit-Stub, Tiers and Waxman, with the difference that every end host randomly select a router to attach to. Consistent with AS-level topologies, delay ratio (Fig. 19) and link stress ratio (Fig. 20) are closely related to the underlying network topology models, while cost ratio and control overhead ratio are not. Overall, among the three topology generators, Transit-Stub matches the AT&T backbone best, followed by Tiers and Waxman.
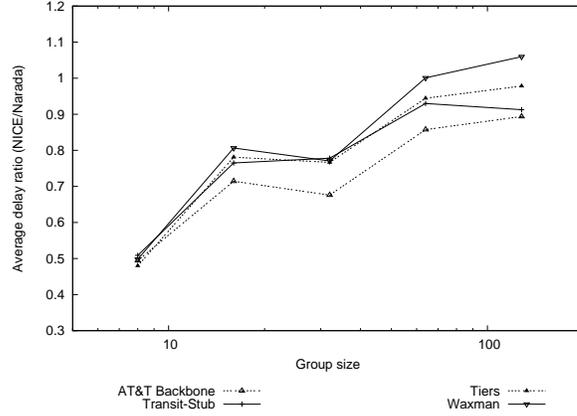
23

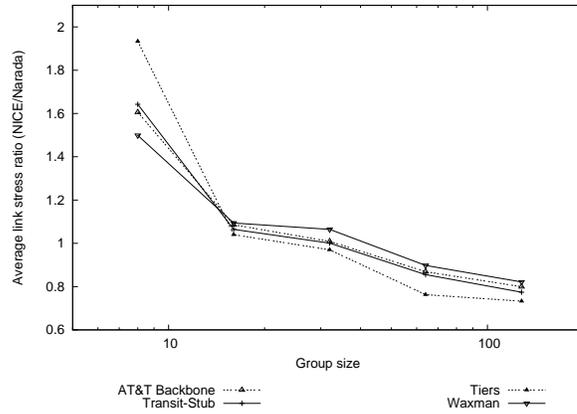Fig. 19. Delay ratio of NICE/Narada for router-level graphs.

Fig. 20. Link stress ratio of NICE/Narada for router-level graphs.

4.5 Lessons Learned

We demonstrate through simulations that different graph models could produce strong deviation in multicast performance evaluation. For example, delay ratio and efficiency of multicast routing algorithms, and link stress ratio of application layer multicast are affected significantly by different topologies, whereas cost ratio does not seem to be so sensitive to topologies. In addition, we observe that even with the same topology generator, different parameter settings drastically change the multicast performance (which is not shown due to space limitation). Therefore, it is critical for researchers to select topology models and parameters that best resemble the real network in terms of multicast performance.

Simulation results of multicast efficiency and multicast state aggregation show that these metrics are sensitive to group properties such as group size and group participation probabilities. Since applications may have different group size and member distributions, it is very important and necessary that the

evaluation of multicast protocols consider these factors and adjust the simulation environment towards target applications.

5   Conclusion

In this paper, we propose a realistic and systematic performance evaluation framework. Our work can be summarized as follows:

- We identify a wide variety of multicast performance metrics.
- We investigate a variety of real and generated network topologies at both AS-level and router-level (especially those produced by power-law topology generators), and examine their impact on the performance of various multicast protocols.
- We propose a generic group participation probability model (truncated normal distribution), which contains the commonly accepted uniform membership model, evaluate multicast performance with this membership model and explore how different parameters affect various multicast metrics.
- We conduct three case studies, namely, multicast routing protocols CBT vs. SBT, multicast state scalability, and application layer multicast, for which we focus on the topology sensitivity and group distribution sensitivity of multicast performance metrics.

Through thorough analysis and extensive simulation studies, we draw the following conclusions:

- Unrealistic topology models can significantly affect multicast performance of various protocols (including application layer multicast).
- We identify the GLP topology generator (for AS-level topology) and the Transit-Stub topology generator (for router-level topology) to be realistic in terms of multicast performance metrics that we investigated.
- Many multicast performance metrics, such as delay ratio, multicast efficiency, and state scalability are sensitive to group size.
- For protocols/approaches exploiting multiple group properties (such as inter-group aggregation), their performance can be significantly affected by the group membership distribution.
- Therefore, simulation studies on multicast should appropriately gear simulation environments with realistic topology models and group models that best match their target applications.

Future Work While we have presented a realistic and systematic multicast evaluation framework and investigated many typical metrics and representative case studies, there are still many issues to be addressed in future work. First, we would incorporate other spatial group properties (such as clustering

25

and group participation correlation) and temporal group properties into our framework. Second, we would like to investigate the impact of small-world property of topology graphs on multicast performance. We also want to use a more detailed network topology model to take link asymmetry, link capacity and delay, and background traffic into consideration. Third, more case studies, such as QoS multicast and reliable multicast, obviously deserve further exploration.

References

[1] CAIDA, the Cooperative Association for Internet Data Analysis. *http://www.caida.org/*.

[2] NLANR Global ISP interconnectivity by AS number. *http://moat.nlanr.net/AS/*.

[3] TIERS. *http://www.isi.edu/haldar/topogen/tiers*.

[4] University of oregon route views project. *http://antc.uoregon.edu/route-views/*.

[5] AT&T IP backbone. *http://www.ipservices.att.com/backbone/*, 2001.

[6] K. C. Almeroth and M. H. Ammar. Collecting and modeling the join/leave behavior of multicast group members in the MBone. In *Proceedings of Higher Performance Distributed Computing*, Aug. 1996.

[7] A. Ballardie. Core Based Trees (CBT version 2) multicast routing: protocol specification. *IETF RFC 2189*, Sept. 1997.

[8] S. Banerjee, C. Kommareddy, and B. Bhattacharjee. Scalable application layer multicast. In *Proceedings of ACM SIGCOMM*, Aug. 2002.

[9] A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, Oct. 1999.

[10] D. Barr and E. Sherrill. Mean and variance of truncated normal distributions. *The American Statistician*, 53:357–361, 1999.

[11] T. Billhartz, J. B. Cain, E. Farrey-Goudreau, D. Fieg, and S. G. Batsell. Performance and resource cost comparisons for the CBT and PIM multicast routing protocols. *IEEE Journal on Selected Areas in Communications*, 15(3):304–315, 1997.

[12] T. Bu and D. Towsley. On distinguishing between Internet power law topology generators. In *Proceedings of IEEE INFOCOM*, June 2002.

[13] K. Calvert, E. Zegura, and S. Bhattacharjee. How to model and internetwork. In *Proceedings of IEEE INFOCOM*, Mar. 1996.

[14] R. Chalmers and K. Almeroth. Modeling the branching characteristics and efficiency gains of global multicast trees. *Proceedings of IEEE INFOCOM*, Apr. 2001.

[15] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. The origin of power laws in Internet topologies revisited. In *Proceedings of IEEE INFOCOM*, June 2002.

[16] S. Chen, K. Nahrstedt, and Y. Shavitt. A QoS-aware multicast routing protocol. *Proceedings of IEEE INFOCOM*, Mar. 2000.

[17] Y.-H. Chu, S. G. Rao, S. Seshan, and H. Zhang. Enabling conferencing applications on the Internet using an overlay multicast architecture. In *Proceedings of ACM SIGCOMM*, Aug. 2001.

[18] Y.-H. Chu, S. G. Rao, and H. Zhang. A case for end system multicast. In *Proceedings of ACM SIGMETRICS*, June 2000.

[19] J. Chuang and M. Sibru. Pricing multicast communications: A cost-based approach. *Proceedings of the Internet Society INET'98 Conference*, July 1998.

[20] J.-H. Cui, M. Faloutsos, D. Maggiorini, M. Gerla, and K. Boussetta. Measuring and modelling the group membership in the Internet. In *Proceedings of the ACMSIGCOMM/USENIX Internet Measurement Conference (IMC 2003)*, Oct. 2003.

[21] S. Deering, D. Estrin, D. Farinacci, and V. Jacobson. Protocol Independent Multicast (PIM), Dense Mode Protocol : Specification. *Internet draft*, Mar. 1994.

[22] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, and L. Wei. Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification. *IETF RFC 2362*, June 1998.

[23] S. Fahmy and M. Kwon. Characterizing overlay multicast networks. In *Proceedings of IEEE ICNP*, Nov. 2003.

[24] M. Faloutsos, A. Banerjea, and R. Pankaj. QoSMIC: Quality of Service sensitive Multicast Internet protoCol. *Proceedings of ACM SIGCOMM*, Sept. 1998.

[25] M. Faloutsous, P. Faloutsous, and C. Faloutsous. On power-law relationships of the Internet topology. In *Proceedings of ACM SIGCOMM*, Sept. 1999.

[26] A. Fei, J.-H. Cui, M. Gerla, and M. Faloutsos. Aggregated Multicast: an approach to reduce multicast state. *Proceedings of Sixth Global Internet Symposium (GI2001)*, Nov. 2001.

[27] A. Fei, J.-H. Cui, M. Gerla, and M. Faloutsos. Aggregated Multicast with inter-group tree sharing. *Proceedings of NGC*, Nov. 2001.

[28] A. Fei and M. Gerla. Receiver-initiated multicasting with multiple QoS constraints. *Proceedings of IEEE INFOCOM*, Mar. 2000.

[29] P. Francis. Yoid: extending the internet multicast architecture. *http://www.aciri.org/yoid/docs/index.html*.

[30] T. Henderson and S. Bhatti. Modelling user behaviour in networked games. In *Proceedings of the 9th ACM Multimedia Conference*, Oct. 2001.

[31] J. Hou, H.-Y. Tyan, B. Wang, and Y.-M. Chen. QoS extension to CBT. *Internet draft: draft-hou-cbt-qos-00.txt*, Feb. 1999.

[32] C. Jin, Q. Chen, and S. Jamin. Inet: Internet topology generator. Technical report, UM CSE-TR-433-00, 2000.

[33] S. Jin and A. Bestavros. Small-world characteristics of Internet topologies and multicast scaling. In *Proceedings of 11th International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2003)*, Oct. 2003.

[34] A. Medina, I. Matta, and J. Byers. On the origin of power laws in Internet topologies. *ACM SIGCOMM Computer Communication Review*, 30(2):18–34, Apr. 2000.

[35] P. V. Mieghem, G.Hooghiemstra, and R. van der Hofstad. On the efficiency of multicast. *IEEE/ACM Transactions on Networking*, Dec. 2001.

[36] J. Moy. Multicast routing extensions to OSPF. *RFC 1584*, Mar. 1994.

[37] C. Partridge, D. Waitzman, and S. Deering. Distance Vector Multicast Routing Protocol. *RFC 1075*, 1988.

[38] D. Pendarakis, S. Shi, D. Verma, and M. Waldvogel. ALMI: An application level multicast infrastructure. In *Proceedings of the 3rd USNIX Symposium on Internet Technologies and Systems*, Mar. 2001.

[39] G. Philips and S. Shenker. Scaling of multicast trees: comments on the Chuang-Sirbu scaling law. In *Proceedings of ACM SIGCOMM*, Sept. 1999.

[40] P. Radoslavov, H. Tangmunarunkit, H. Yu, R. Govindan, S. Shenker, and D. Estrin. On characterizing network toplogies and analyzing their impact on protocol design. Technical report, USC Dept. of CS, USC-CS-TR-00-731, 2000.

[41] P. I. Radoslavov, D. Estrin, and R. Govindan. Exploiting the bandwidth-memory tradeoff in multicast state aggregation. Technical report, USC Dept. of CS, 99-697 (Second Revision), July 1999.

[42] I. Stoica, T. Ng, and H. Zhang. REUNITE: A recursive unicast approach to multicast. In *Proceedings of IEEE INFOCOM*, Tel Aviv, Israel, Mar. 2000.

[43] H. Tangmunarunkit, R. Govindan, and S. Jamin. Network topology generators: Degree-based vs. structual. In *Proceedings of ACM SIGCOMM*, Aug. 2002.

[44] D. Thaler and M. Handley. On the aggregatability of multicast forwarding state. *Proceedings of IEEE INFOCOM*, Mar. 2000.

[45] J. Tian and G. Neufeld. Forwarding state reduction for sparse mode multicast communications. *Proceedings of IEEE INFOCOM*, Mar. 1998.

[46] B. M. Waxman. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, 6(9):1617–1622, Dec. 1988.

[47] L. Wei and D. Estrin. The trade-offs of multicast trees and algorithms. *Proceedings of the International Conference on Computer Communications and Networks (ICCCN)*, 1994.

[48] T. Wong and R. Katz. An analysis of multicast forwarding state scalability. In *Proceedings of the 8th International Conference on Network Protocols (ICNP)*, Japan, Nov. 2000.

[49] E. W. Zegura, K. L. Calvert, and M. J. Donahoo. A quantitative comparison of graph-based models for Internet topology. *IEEE/ACM Transactions on Networking*, 5(6):770–783, Dec. 1997.