Characterizing Traffic Demand Aware Overlay Routing Network Topologies

Benjamin D. McBride
Kansas State University
2061 Rathbone Hall
Manhattan, KS 66506
Email: bdm7935@ksu.edu

Caterina Scoglio
Kansas State University
2061 Rathbone Hall
Manhattan, KS 66506
Email: caterina@ksu.edu

***Abstract*— We introduce a traffic demand aware cost model for the creation of overlay routing networks. We investigate the effects on the created network topologies when the traffic demand between nodes is considered. The resulting network topologies often differ dramatically from topologies formed when traffic demand between nodes is not weighed. We found that the global network characteristics were changed as well as individual node placement in the topology. Our results clearly show that the overlay network topology is adapted to reflect the underlying traffic demand distribution. We conclude that overlay network topologies formed using a traffic demand aware cost model are better suited to carry the traffic demand between nodes.**

I. Introduction

The growth of peer-to-peer and multimedia applications has created a greater need for performance and quality of service (QoS) guarantees over the Internet. Unfortunately, the Internet does not provide such guarantees. Overlay routing networks, or simply, overlay networks, have been presented as a solution to these functional limitations of the Internet [1] [2] [3]. An overlay network is an application-layer, logical network. Applications can route data through the underlay network directly, or through intermediate nodes in the overlay network. Consequently, applications can choose routing paths that provide improved performance and QoS based on a variety of metrics. An open problem in the development of overlay networks is the creation of the network topology.

Game-theoretic approaches to network topology creation have been proposed [4] [5]. These approaches define a cost model, and nodes act selfishly to create logical links in the overlay network such that their cost is minimized. While both approaches consider network locality, neither considers the traffic demand between nodes in their cost models. We believe that network topologies created using a traffic demand aware cost model will differ significantly from topologies formed by traffic demand ignorant cost models. We further believe that the resulting topologies will be reflective of the underlying traffic demand patterns.

We propose a cost model that includes the traffic demand between nodes. This is an extension of the cost models presented in previous work [4] [5] and is suggested by Fabrikant, et al. in Section 6 [4]. A closely related cost model has been presented by the authors [6]. We include a thorough analysis of the topologies that are formed with the traffic demand aware cost model. During our analysis, we introduce the idea of *popular* and *greedy* nodes to characterize common traffic demand distributions. We find traffic demand aware overlay networks often differ dramatically with respect to the overall topology and node placement when different traffic demand distributions are considered. Our results clearly show that network topologies formed from traffic demand aware cost models are adapted to reflect the underlying traffic demand distribution and carry the traffic demand effectively.

This work differs from the work previous work [4] [5] by explicitly examining the effect of traffic demand on the overlay network topologies formed. Other cost models have been proposed that consider traffic demand and load between nodes [7]. However, they only analyze the cost of existing overlay network topologies, rather than considering the effect on network formation.

The remainder of this paper is as follows. In Section II we examine related work. We introduce a traffic demand aware cost model for network formation in Section III. Section IV presents the effects of considering traffic demand on overlay network topologies. We conlude and discuss future work in Section V.

II. Related work

Most overlay network topologies can be classified as either structured or unstructured. Structured overlay networks form organized topologies and then determine routing paths using common network routing protocols. RON is built on a complete graph topology [1]. Several systems, such as CAN [8] and Tapestry [2], use dynamic hash tables to form the overlay mesh. While most structured network topologies are static, some approaches, such as Narada [9], adapt the topology to changing network dynamics. Unstructured overlay networks topologies select neighbors through a predominately random process. Most modern peer-to-peer systems, such as Gnutella, are unstructured [10].

Recently, game-theoretic approaches have been proposed to create overlay network topologies [4] [5] [11]. Nodes pay for the logical links they establish and benefit from the network formed. Nodes are selfish, acting in their own interest to minimize their cost. Our work is an extension of the cost model presented by Fabrikant, et al. [4]. Others have used this cost model to characterize selfishly constructed overlay networks [5]. However, previous research has not considered the impact of traffic demand on the overlay network topologies.

The weighted network creation game, with weights corresponding to the traffic demand between nodes, is equivalent to our cost model when the linking cost is equal to the number of links created. Upper and lower bounds on the price of anarchy for the weighted network creation game are given by Albers, et al. [12]. Using the stretch instead of the distance between nodes has been proposed [11]. Related cost models have also been proposed in which nodes share the cost of edges and can pay for non-adjacent edges [13].

III. Traffic Demand Aware Cost Model

We assume that each node needs to select its neighbors in a distributed fashion. Each node has imperfect information and does not know the other nodes' neighbors. Let $G = (N, L)$ be the graph representing the overlay network and $G_u = (N, E)$ be the graph representing the underlay, or physical, network. N is the set of nodes that are in both the overlay and physical network, while the set of logical links L can be different from the set of physical links E. A logical link $l \in L$ is constructed on a path composed of physical links $e \in E$. Each node $i \in N$ has a traffic demand toward a node subset $S_i \subseteq N$. Let $t_{i,j}$ be the traffic demand between node i and node j in the subset S_i. The objective for each node is to create logical links to a subset of nodes, $B_i \in 2^{N-\{i\}}$, such that its total cost is minimized. We define cost using two components:

1) Link cost: cost to create and maintain a logical link between node i to node j
2) Transport cost: cost to transport the traffic demand between nodes i and j

The cost for node i to connect to each node $j \in B_i$ and carry traffic demand $t_{i,j}$ to each node $j \in S_i$ is defined as the sum of the link cost and transport cost:

$$C_i = \alpha \sum_{j \in B_i} h_{i,j} + \sum_{j \in S_i} d_G(i,j) t_{i,j} \tag{1}$$

where α is the relative cost of creating a logical link to the cost of transporting the traffic demand through the existing network, $h_{i,j}$ is the linking cost between i and j, and $d_G(i,j)$ is the distance in the overlay network to node j (∞ if j is unreachable from i). The total cost of the overlay network is defined as:

$$C(G) = \sum_{i \in N} C_i \tag{2}$$

It is important to note that the linking cost between i and j, is a general function that can represent a wide variety of metrics. The transport cost term in (1) can also be thought of as a generalized distance function [5]. Additionally, once a logical link has been established from i to j any node in the network can use the link. We do not consider the link to be directed in terms of its use and for calculation of d_G.

IV. Results

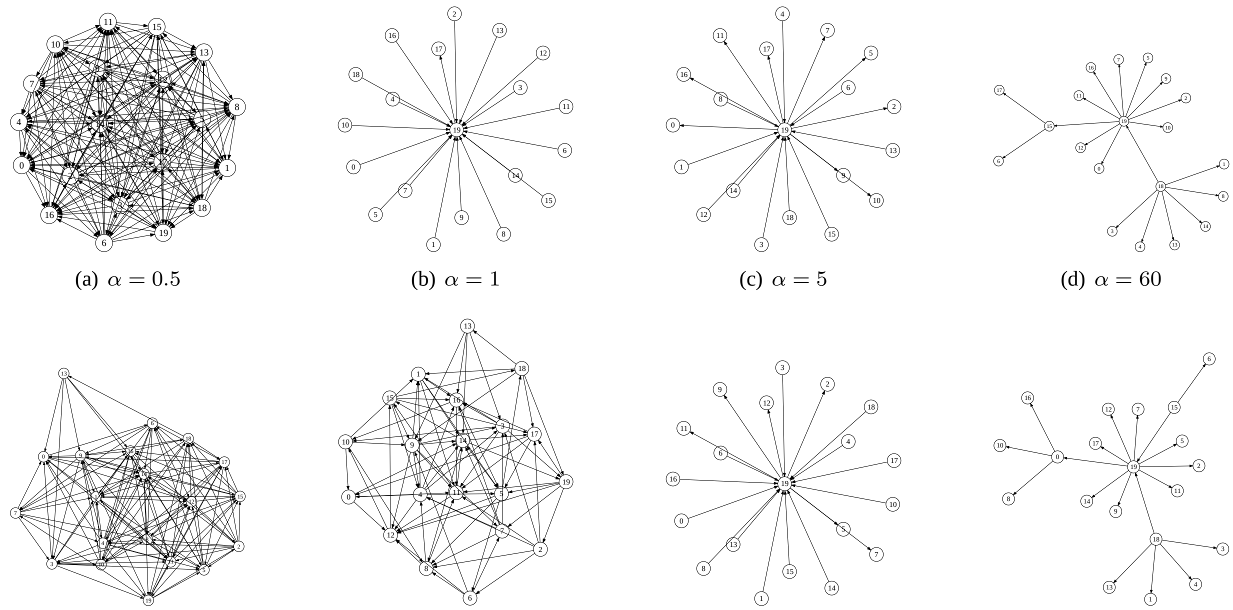
We start with a randomly connected overlay graph, $|N| = 20$, and assume a fully connected underlay with $h_{i,j} = 1$ unless specified otherwise. The traffic demand is normally distributed with mean and variance of one. Negative traffic demand values are rounded to zero. Like previous work [5], we use an iterative exhaustive search of each node's *strategy* space until equilibrium is reached. Each node, in turn, chooses a strategy, $B_i \in 2^{N-\{i\}}$, that minimizes its cost with respect to the rest of the graph. This process is repeated until a Nash equilibrium is reached, that is, no node can lower its cost without some other node first increasing its cost. An equilibrium is reached using this procedure, but it is not necessarily a global optimal network [4] [12].

A. Normal traffic demand

We first contrasted the overlay network topologies formed when the traffic demand between nodes is not considered with those formed when normal traffic demand between nodes is considered. The iterative exhaustive search procedure was used until equilibrium was reached. Figure 1 shows the resulting topologies for differing values for α.

For $\alpha = 0.5$, the network topology is a partially connected graph when traffic demand is considered (Figure 1(e)). The change in topology increases the transport cost slightly, but the link cost is greatly reduced over the complete graph topology formed when traffic demand is ignored (Figure 1(a)). On the other hand, when $\alpha = 1$, as in Figures 1(b) and 1(f), the link cost is higher when traffic demand is considered, but the transport cost is considerably lower. It should be noted that our results differ from the work of Chun et al. [5] with respect to $\alpha = 1$. We found the equilibrium topology to be a star rather than a partially connected graph when the traffic demand between nodes is ignored. However, it is noted by Fabrikant et al. [4] that the star topology is also a Nash equilibrium for $\alpha = 1$, albeit the worst one. When $\alpha = 5$, there is no change in the cost when considering traffic demand as both topologies are stars, Figures 1(c) and 1(g). Surprisingly, the total network cost is actually slightly higher when $\alpha = 60$ and normal traffic demand is used, Figures 1(d) and 1(h).

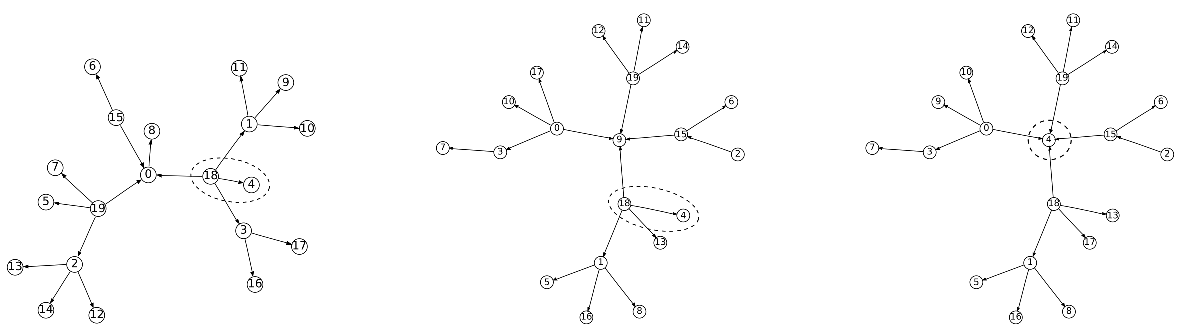
We next constrained the maximum degree of a node by setting the linking cost $h_{i,j}$ to infinity if either node i or j has degree equal to the maximum degree [5]. We assume a maximum degree of four. Figure 2 shows the network topologies for varying traffic demand models when $\alpha = 60$. Figures 2(a) and 2(b) show the topologies formed when traffic demand is not considered and when normal traffic demand is considered, respectively. Both trees have a diameter of six and have equivalent node degree distributions. However, the characteristic path length is decreased when the traffic demand is considered. For $\alpha <= 5$, there was little change in the overall topologies formed when considering normal traffic demands. The changes that occurred were more subtle. We expound on these types of changes in the following sections.



(a) $\alpha = 0.5$ (b) $\alpha = 1$ (c) $\alpha = 5$ (d) $\alpha = 60$

(e) $\alpha = 0.5$ (f) $\alpha = 1$ (g) $\alpha = 5$ (h) $\alpha = 60$

Fig. 1. Equilibrium network topologies found using iterative exhaustive search. The cost model used for graphs (a)-(d) does not account for traffic demand between nodes. Graphs (e)-(h) were created using a traffic demand aware cost model. The direction of the edge between nodes indicates which node paid for the logical link.



(a) Traffic demand ignorant (b) Normal traffic demand (c) Node 4 is popular

Fig. 2. Network topologies for different traffic demand models when the maximum node degree is constrained and $\alpha = 60$. The traffic demand ignorant network in Figure (a) has a higher characteristic path length than the traffic demand aware network in Figure (b) despite the networks having equivalent diameter and node degree distribution. Node 4 is a root node, rather than a leaf, when the traffic demand towards it is increased, Figure (c).

B. Popular nodes

We next considered the concept of *popular* nodes. A popular node is a node that has traffic demand towards it that is significantly higher than the mean traffic demand between nodes. We expected the presence of popular nodes would have a large impact on the topologies created. We also expected the basic topologies would remain the same, but the popular nodes would become root nodes rather than leaf nodes in the tree topologies. We used the normal traffic demand pattern described previously, but with an increase of the mean traffic demand towards node 4 by a factor of five. Initially, we did not constrain the degree of the nodes.

As expected the overall network topologies remained similar when using the modified traffic demand distribution. However, as shown in Table I, the degree of node 4 has increased for $\alpha < 60$. The change in degree is a direct result of other nodes creating logical links towards node 4, as evident when observing the in degree of node 4. The implications of this change are clearly seen in Figures 3(a) and 3(b). When node

TABLE I

DEGREE OF POPULAR NODE FOR VARYING VALUES OF α

	Normal traffic demand		Node 4 is popular	
α	degree	in degree	degree	in degree
0.5	16	6	19	9
1	12	5	19	12
5	1	0	19	19
60	1	1	1	1

4 is popular it becomes the root node in the topology rather than a leaf node.

Surprisingly, when $\alpha = 60$, we see no change in the topology. Node 4 is still a leaf node. While this result was initially surprising, on closer examination it is easily explained. Observe in Figure 4(a) that most of the nodes do not create any logical links in the overlay network. These nodes use the network for their benefit without contributing to the rest of the network. This freeriding behavior in overlay networks has been observed [5] and extensivly studied by others [14].

While there is no change in the topology for $\alpha = 60$ when the maximum node degree is unconstrained, the topology does change when the maximum node degree is constrained. Figures 2(b) and 2(c) show the result of node 4 being popular. Under normal traffic demand conditions, node 4 is a leaf node. However, when node 4 is popular it becomes a root node in the topology. This happens because fewer nodes are able to freeride when the maximum node degree is constrained. This results in a topology that is better adapted to the underlying traffic demand distribution.

C. Greedy nodes

In addition to popular nodes, we examined the effect of *greedy* nodes on the network topology. A greedy node is a node that has significantly higher traffic demand towards other nodes than the mean traffic demand between nodes. We expected greedy nodes to take advantage of the existing network topology to decrease transport costs.

We used the normal traffic demand distribution described previously, with an increase by a factor of five in the mean traffic demand of node 4 towards other nodes. We do not constrain the degree of the nodes.

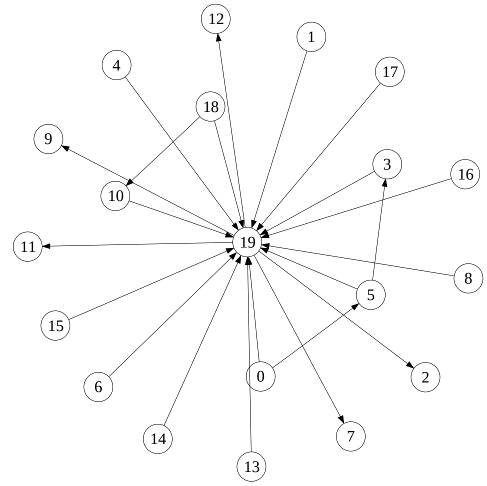
TABLE II

DEGREE OF GREEDY NODE FOR VARYING VALUES OF α

	Normal Traffic Demand		Node 4 is "greedy"	
α	degree	out degree	degree	out degree
0.5	16	10	19	13
1	12	7	19	14
5	1	1	19	16
60	1	0	1	1

As expected the overall network topologies remained similar. However, as shown in Table II, the degree and out degree of node 4 significantly increased. For $\alpha = 5$, node 4 has become the root node in the star topology rather than a leaf node, see Figure 3. This is a result of node 4 creating more logical links towards the other nodes as shown by the increase in the out degree. Another interesting result is when $\alpha = 60$. As noted previously, when α is large most nodes have little incentive to create logical links towards other nodes. But when node 4 is greedy, it exploits the existing network topology and creates a logical link towards the root node of the tree, as shown in Figure 4.

Fig. 5. Network topology for $\alpha = 5$. Nodes 0, 5, and 18 have high traffic demand towards nodes 5, 3, and 10, respectively, and therefore, create additional logical links towards them when traffic demands are considered in the cost model.



D. High traffic demand links

We next considered increased traffic demand between just two nodes. This differs from the popular and greedy nodes in that the traffic demand remains the same towards the most other nodes; there is just a specific node towards which there is high traffic demand. Figure 5 shows the network topology for such a scenerio. Surprisingly, the graph is no longer a tree. The basic star topology can clearly be identified, but additional logical links have formed where there is high traffic demand between the nodes.

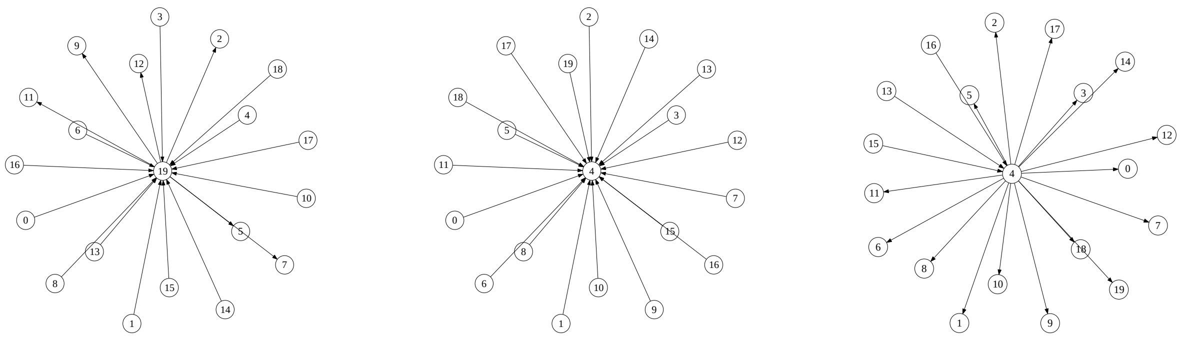
V. CONCLUSIONS AND FUTURE WORK

We introduced a cost model for overlay network creation that includes the traffic demand between nodes. We showed that the traffic demand aware cost model formed network topologies with reduced link cost and transport cost, depending on the network parameters. We also introduced the concept of popular and greedy nodes and showed the effects of such nodes on the network topologies. Finally, we showed that high traffic demand links have a profound effect on the network topology. Consequently, we conclude that overlay network topologies formed from a traffic demand aware cost model are better suited to carry the traffic demand between nodes.

For future work we will consider networks of larger size and realistic topologies. We also plan to study the effects of heterogeneous nodes in the formation of overlay network topologies. This will necessitate the use of heuristic and stochastic approaches to network formation, since the exhaustive search method is exponential time complexity in the number of nodes. We plan to use the network characteristics found in this work to develop the heuristic approaches.

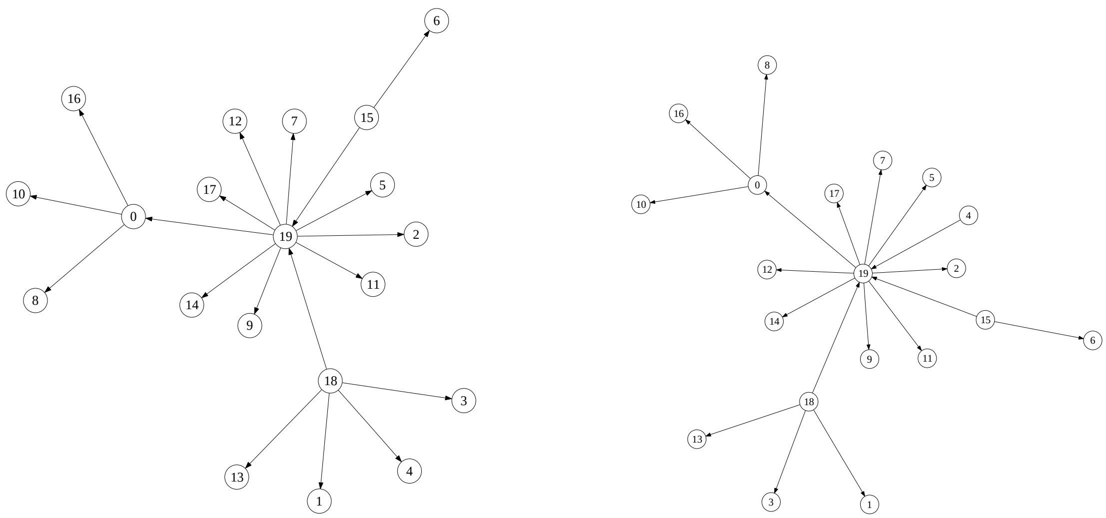
REFERENCES

[1] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris, "Resilient overlay networks," in *SOSP 2001: Proceedings of the Eighteenth ACM Symposium on Operating Systems Principles*, 2001.



(a) Normal traffic demand (b) Node 4 is popular (c) Node 4 is greedy

Fig. 3. Network topologies for different traffic demand models for $\alpha = 5$. Figure (a) shows the network topology when normal (gaussian) traffic demands are considered. Figures (b) and (c) show the network topology when node 4 is popular and greedy, respectively. When node 4 is popular, that is, other nodes have a high traffic demand towards it, it becomes the root node by virtue of the other nodes creating logical links towards it. Conversely, when node 4 has a high traffic demand towards other nodes, it becomes the root node by virtue of it creating logical links towards the other nodes.



(a) Normal traffic demand (b) Node 4 is greedy

Fig. 4. Network topology comparison between normal traffic demand and greedy node traffic demand for $\alpha = 60$. Figure (a) shows the network topology when there is normal traffic demand. Figure (b) shows the network topology when node 4 is greedy. Notice that node 4 is a leaf node that creates no logical links toward other nodes in Figure (a), but when node 4 is greedy, that is, it has a high traffic demand towards other nodes, it exploits the existing network topology and creates a logical link towards the root node.

[2] B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. D. Kubiatowicz, "Tapestry: A resilient global-scale overlay for service deployment," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 1, 2004.

[3] H. Zhang, J. F. Kurose, and D. Towsley, "Can an overlay compensate for a careless underlay?" in *INFOCOM 2006: Twenty-fifth Annual Joint Conference of the IEEE Computer and Communications Societies*, 2006.

[4] A. Fabrikant, A. Luthra, E. Maneva, C. H. Papadimitriou, and S. Shenker, "On a network creation game," in *PODC 2003: Proceedings of the Twenty-second Annual Symposium on Principles of Distributed Computing*, 2003.

[5] B.-G. Chun, R. Fonseca, I. Stoica, and J. Kubiatowicz, "Characterizing selfishly constructed overlay routing networks," in *INFOCOM 2004: Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, 2004.

[6] B. McBride, C. Scoglio, and S. Das, "Distributed biobjective ant colony algorithm for low cost overlay network routing," in *Proceedings of the 2006 International Conference on Artifical Intelligence*, 2006.

[7] N. Christin and J. Chuang, "A cost-based analysis of overlay routing geometries," in *INFOCOM 2005: Twenty-fourth Annual Joint Conference of the IEEE Computer and Communications Societies*, 2005.

[8] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network," in *Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2001.

[9] Y.-H. Chu, S. G. Rao, and H. Zhang, "A case for end system multicast," in *Measurement and Modeling of Computer Systems*, 2000.

[10] D. Stutzbach, R. Rejaie, and S. Sen, "Characterizing unstructured overlay topologies in modern P2P file-sharing systems," in *Proceedings of Internet Measurement Conference 2005*, 2005.

[11] T. Moscibroda, S. Schmid, and R. Wattenhofer, "On the topologies formed by selfish peers," in *Proceedings of the Twenty-fifth Annual ACM Symposium on Principles of Distributed Computing*, 2006.

[12] S. Albers, S. Eilts, E. Even-Dar, Y. Mansour, and L. Roditty, "On Nash equilibria for a network creation game," in *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2006.

[13] E. Anshelevich, A. Dasgupta, E. Tardos, and T. Wexler, "Near-optimal network design with selfish agents," in *Thirty-fifth Annual ACM Symposium on Theory of Computing*, 2003.

[14] B. G. Rocha, V. Almeida, and D. Guedes, "Increasing QoS in selfish overlay networks," *IEEE Internet Computing*, vol. 10, no. 3, 2006.