Performance study of peer-to-peer video streaming on complex networks

Lyrial Chism*, Xiaoqing He**, Liquan Huang***, Ashraf Ibrahim†
Christopher Jones††, Yan Shu‡, Chai Wah Wu†‡, Zhenghua Fu†‡
*University of Mississippi, **University of Minnesota, ***University of Delaware †Texas A&M University
††University of Pittsburgh, ‡Georgia Institute of Technology, †‡IBM T. J. Watson Research Center
contact email: chaiwahwu@ieee.org

***Abstract*—In this paper we study the video streaming bandwidth of peer-to-peer streaming networks where the underlying topology is a complex network. We focus on the maximal streaming rate and how it depends on the type of network. We consider networks such as small world networks, scale free networks, locally connected networks and random networks. The experimental results indicate that a more connected graph does not necessarily imply a higher streaming rate, whereas properties such as the existence of a Hamiltonian path from the source does.**

I. Introduction

Recently, peer-to-peer based data transmission has emerged as a promising technology for distributing data to users at internet scale. Many applications such as BitTorrent and EMule have successfully utilized the internet to distributes files to thousands of users. In a peer-to-peer model, in comparison to a client-server model, each peer can both transmit and receive data between each other. The main advantage of a peer-to-peer architecture is potentially a much smaller server load than the client-server model, especially in a large scale network. It also scales well with the number of clients.

The possibility of using peer-to-peer networks for video streaming and video-on-demand has recently been studied [7]. The current client-server model of online video streaming and distribution incurs a significant bandwidth cost for content providers and the peer-to-peer framework promises to reduce significantly the bandwidth needed from the source. Unlike peer-to-peer file sharing, peer-to-peer video streaming has additional constraints and significant challenges in its implementation. One of the most important constraints is imposed by real-time streaming. Users are expected to watch a downloaded video immediately (or with a few seconds of delay). This means that the data must be downloaded quickly in an ordered way. With applications such as BitTorrent, files may take several hours and sometimes days to download and the order in which the data packets are downloaded is not important since the file is not used until it is complete.

This paper focuses on characterizing the maximal streaming rate of data from a single source to its peers in a given network. We analyzing different types of complex networks used to model communication networks and they affect the streaming rate.

II. Maximum Achievable Streaming Rate Problem

For a given network, we want to find the maximum achievable streaming rate. The maximum achievable rate is defined as the maximum streaming rate r^* at which the source can deliver content to all the nodes [4]. We model the network as a graph $G = \{V, E\}$ of n nodes and an upload bandwidth u_i for each node i. We define node s to be the source node (server) and all other nodes to be *peers*. Since users are able to download and upload data at different rates, our graphs are directed. We assume that the download bandwidth of each peer is not a bottleneck and can be assumed to be infinite. $e(i,j)$ is an edge of the graph if data can travel from node i to node j at some rate r that is no more than u_i. An upper bound for the streaming rate in any bandwidth scenario is given by $\min\left(u_s, \frac{\sum_{i\in N} u_i}{n-1}\right)$ [6]. This rate is achievable for the complete graph[1].

We define a packet spanning tree as a directed spanning tree rooted at the source node. It has been shown (see e.g. [4]) that we can attain the maximum achievable rate by finding the maximal number of packet spanning trees that can be packed in the graph. It is easy to show that finding the maximum achievable rate is a NP-hard problem. For a given connected graph, if the bandwidth of each node is 1, then the maximum achievable rate is either 0 or 1. Finding the maximum achievable rate in this graph is then equivalent to finding a Hamiltonian path starting from the source. In this case the maximum achievable rate is 1 if and only if there is a Hamiltonian path starting from the source. Since finding a Hamiltonian path in a connected graph is NP-hard problem, finding the maximum achievable rate is a NP-hard problem. This argument also shows that if all the bandwidth u_i are equal, then the upper bound shown above is equal to u_s and this bound is achievable if there is a Hamiltonian path starting from the source.

Due to the intractability of the maximum achievable rate problem, Ref. [4] gives an linear programming (LP) relaxation of the maximum achievable rate problem. The approach is to find a distribution of bandwidth among the links that satisfy the constraints, while maximizing the minimum capacity from the source to all other nodes which by Edmond's theorem is

[1] Assuming r^* is an integer.

equal to the achievable rate. Let x_e be a flow value on edge $e \in E$ and $f_{i,e}$ be the amount of flow node i gets from the source via edge e. Let R denote the streaming rate. The LP problem is given by:

$$
\begin{aligned}
&\text{maxmize} && R \\
\sum_{e(s,j)\in E} f_{i,e} - \sum_{e(j,s)\in E} f_{i,e} &= && R \quad \forall i \in N \setminus \{s\} \\
\sum_{e(j,i)\in E} f_{i,e} - \sum_{e(i,a)\in E} f_{i,e} &= && R \quad \forall i \in N \setminus \{s\} \\
\sum_{e(b,a)\in E} f_{i,e} - \sum_{e(a,c)\in E} f_{i,e} &= && R \quad \forall i \in N \setminus \{s\}, \\
& && \forall a \in N \setminus \{i, s\} \\
x_e - f_{i,e} &\geq && 0 \quad \forall e \in E, i \in N \setminus \{s\} \\
\sum_{e(i,j)} x_e &\leq && u_i \quad \forall e \in N \\
f_{i,e} &\geq && 0 \quad \forall e \in E, i \in N \setminus \{s\}
\end{aligned}
$$

In our study, $f_{i,e} = 0, \quad \forall e(j,s) \in E$ since we assume that the source does not download from the peers. The optimal value of the above linear programming problem is a good approximation of the maximum achievable rate for their difference is at most 2 [2]. In the sequel we will denote the optimal value of the LP program as $R_{LP}^{\max}$.

We solve this LP problem using the GNU linear programming kit (GLPK) for various types of networks such as random graphs, scalefree graphs, small world graphs and grid graphs and compare the corresponding values of $R_{LP}^{\max}$.

1) For random graphs we consider the model $G(n,p)$ [3] with n nodes and each edge occurs with probability p.
2) For small world graphs we consider the Watts-Strogatz model where random edges are added to a cycle of n nodes.
3) For scale-free graphs we use the iterative preferential attachment model introduced by Barabasi [1]. We start with a complete graph of k nodes and add a new node and k new edges at each iteration. The k edges connect the new node to existing nodes with the k largest degrees. It has been shown that adhoc Peer-to-peer networks can exhibit behaviors of scale free graphs [9].
4) We will also consider m-dimensional grid graphs which are formed by an m-fold graph Cartesian product of the path graph P_n [10].

Using the 4 types of graphs discussed above as models for peer-to-peer network, we consider the following 3 bandwidth scenarios:

1) All peers and source have the same upload bandwidth ($u_i = 10$), we shall refer to this scenario as the *homogeneous* scenario.
2) All peers have a constant upload bandwidth ($u_i = 10$) and the source has an upload bandwidth 100 times as large ($u_s = 1000$), we refer to this scenario as the *strong source* scenario.
3) The peers and source have upload bandwidths taken as samples from a bandwidth distribution of users of a real video streaming application found in [5].

	modem	ISDN	DSL1	DSL2	Cable	Ethernet
Upload	64	256	128	384	768/384	768
share (%)	2.8	4.3	14.3	23.3	18.0	37.3

For the *Cable* upload bandwidth, we used the average of the two upload values given, i.e. 576. We refer to this scenario as the *random bandwidth* scenario.

III. Streaming rates of various graph topologies

A. Random Graphs

1) Random Graphs with Homogeneous Bandwidths: For a fixed number of nodes n, we vary the edge probability p and study the ratio k/c, where k is the average maximum streaming rate of random graphs, and $c = 10$ is the maximal streaming rate (which is achievable by the complete graph which occurs when $p = 1$). We see that the convergence of k/c to 1 as p is increased is faster for larger n as shown by the graph in Figure 1. We believe that this is due to the fact that a random graph contains a Hamiltonian cycle with high probability if $p \geq c\frac{\log n}{n}$ for some constant c [8]. Thus for large n, a smaller probability p is needed to ensure the existence of a Hamiltonian path from the source node.

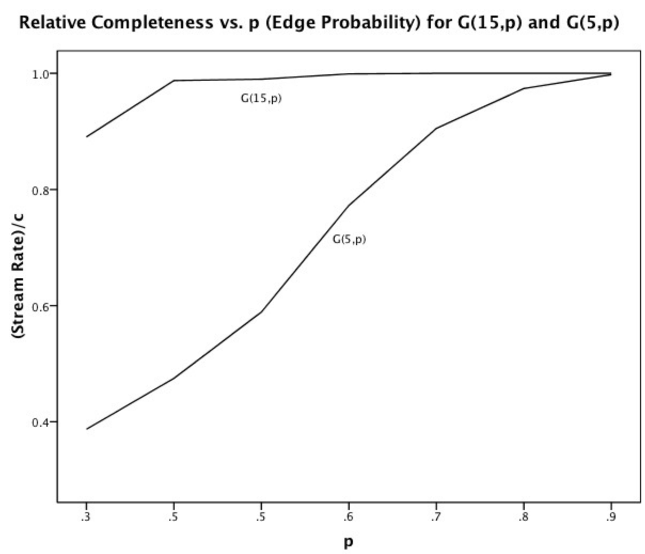

Fig. 1. Average streaming rate of a random graph with respect to edge probability p for $n = 5$ and $n = 15$.

2) Random Graphs with Strong Source Bandwidths: Figure 2 shows the maximum streaming rate distribution for $p = 0.5$ and Figure 3 shows the maximum streaming rate distribution for $p = 0.3$. Notice that in both cases there is a skewness which suggests that we are more likely to obtain a streaming rate higher rather than below the most commonly occuring streaming rate.

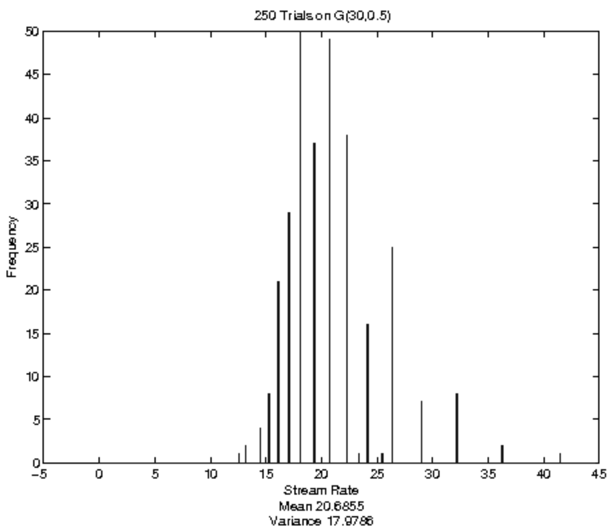

Fig. 2. Distribution of streaming rate for $p = 0.5$ with $u_s = 1000$ and $u_i = 10$.

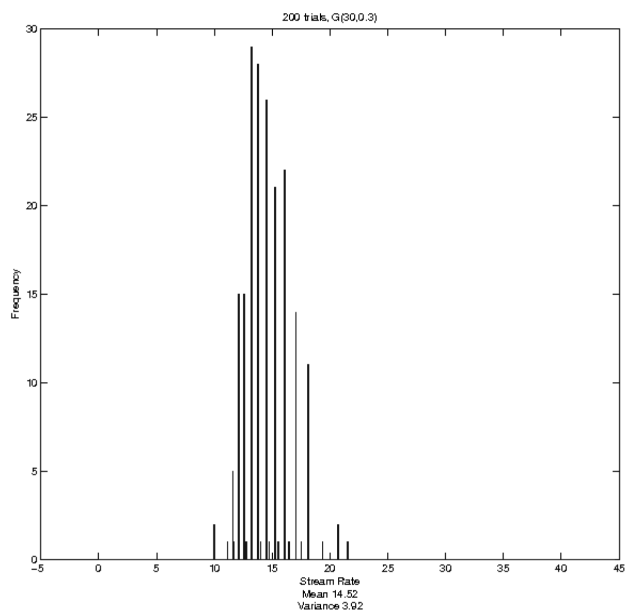

Fig. 3. Distribution of maximum streaming rate of random graph $G(30, 0.3)$

B. Small World Graphs

1) Small World Graphs with Homogeneous Bandwidths: In this model, the maximum streaming rate is always 10. The reason is that the small world graph of this model always contains a cycle and thus a Hamiltonian path from the source, along which the network can achieve its maximum streaming rate.

C. Scale-free Graphs

1) Scale-free Graphs with Homogeneous Bandwidths: For homogeneous bandwidths $u_i = 10$, we generate 300 scale-free graphs and calculate their maximum streaming rates. Figure 4 is a plot of the distribution of these maximum streaming rates. It shows us that over half of the distribution of the streaming rates is at the maximum possible rate 10.

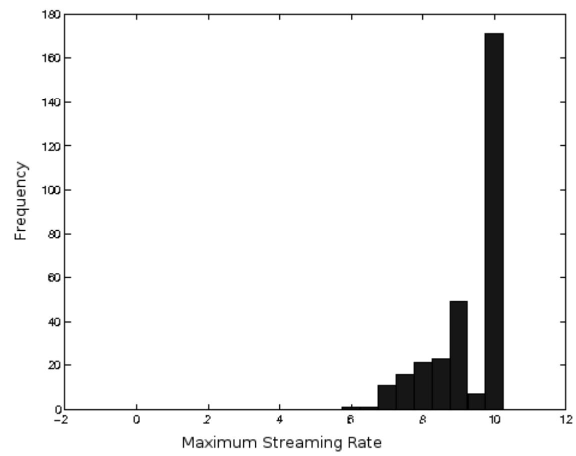

Fig. 4. Distribution of Maximum Streaming Rates for scale free graphs with homogeneous bandwidths.

2) Scale-free Graphs with Strong Source Bandwidths: For strong source bandwidths, we also generate 300 scale-free graphs. The mean of the 300 maximum streaming rates is 10.7318 and the maximum of maximum streaming rates amongst this trial is near 19 (Figure 5).

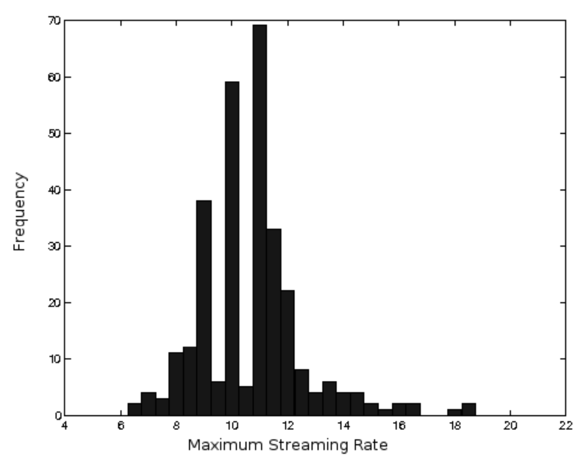

Fig. 5. Distribution of maximum streaming rate of scale free graphs with strong source bandwidths.

3) Scale-free Graphs with Random Bandwidths: For the random bandwidths case, we still generate 300 scale-free graphs. Their maximum streaming rates range from 300 to 600(see Figure 6).

D. Grid Graphs

In this subsection, we focus on 5×5 grid graph. For each node, we calculate the maximum streaming rate of taking this

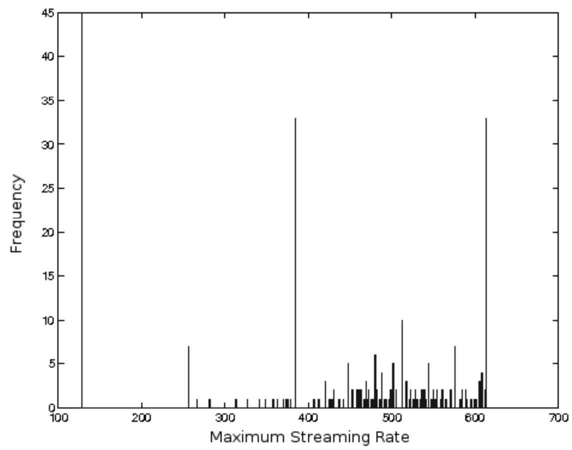

Fig. 6. Distribution of maximum streaming rate of scale free graphs with random bandwidths.

node as source and all the other nodes as peers.

1) Grid Graphs with Homogeneous Bandwidths: In the homogeneous bandwidth case, we find that the maximum stream rates alternate between 9 and 10 when the source node change from one position to its neighborhood. The interesting phenomenon comes from the fact that for any two adjacent nodes in the grid graph, exactly one of them has a Hamiltonian path starting from it.

2) Grid Graphs with Strong Source Bandwidths: We set the bandwidth of the source node as 1000 and the bandwidth of all the other nodes as 10. In the first trial, we take the first node on the boundary as the source. In the second trial, we choose the center node as the source. Surprisingly, in both situation, we obtained the same result as in the homogeneous bandwidths case.

3) Grid Graphs with Random Bandwidths: For randomly generated bandwidths, we generate two samples each of which consists of 100 tests. For each node, we average of the maximum streaming rates of taking this node as source. Just as in the constant bandwidth case, regardless of which node is the source, the average maximum streaming rate does not change much.

The grid graph does not seem to be the ideal model for a peer-to-peer network. Even when the bandwidth at the source is much greater than the bandwidth of the peers, the maximum streaming rate is not large. One noticeable feature of these graphs is that they are stable in the sense that the choice of the source node does not affect the streaming rate.

IV. Discussion

The maximum stream rate are achievable for random graphs which are close to complete, a large percentage of scale-free graphs, Watts-Strogatz small world model and most grid graphs, even though random graphs are considered more connected than small world graphs and grid graphs. This indicate that the connectedness of a graphs is not correlated to the maximum streaming rate. On the other hand, the relatively larger diameter of the small world model and grid graphs implies a larger latency for some peer in the context of video streaming.

V. Future Work

Some questions to answer in future work are the following: Given the dependence of the streaming rate and its robustness on the network topology, how should this guide the way new peers are joined into a peer-to-peer network? For local bandwidth allocation algorithms which are suboptimal, do they exhibit the same dependence on the network topology?

VI. Acknowledgments

This work was the result of the IMA Math. Modeling Workshop for Graduate Students at the University of Minnesota in August 2008. We would like to thank IMA for providing us with this opportunity and such a great working environment. In particular, we thank the workshop organizers, Professors Fadil Santosa and Richard Braun and their staff for making the workshop as stimulating and enjoyable as possible. We also like to thank IBM for giving us the chance to work on this project. We also would like to extend our thanks to the other teams at the workshop and especially their mentors.

References

[1] R. Albert and A. L. Barabasi, "Statistical Mechanics of Complex Networks," *Rev. Mod. Phys.*, vol. 74, pp. 47, 2002.

[2] N.Bansal, R.Khandekar, and V.Nagarajan, "Additive Guarantees for Degree Bounded Directed Network Design", *Proc. ACM STOC*, pp. 769-778, 2008.

[3] B. Bollobas. *Random Graphs*, 2nd Ed., Cambridge University Press, 2001.

[4] Z. Fu. "Throughput Bound in P2P Streaming System in General Overlay Topology." Submitted for publication.

[5] C. Huang, J. Li, and K. Ross. "Can Internet Video-on-Demand be Profitable?" *Proc. ACM SIGCOMM*, pp. 133-144, 2007.

[6] R. Kumar, Y. Liu, and K. Ross. "Stochastic Fluid Theory for P2P Streaming Systems," *Proc. IEEE INFOCOMM*, pp. 919-927, 2007.

[7] B. Li, J. Liu, S. Rao, and H Zhang. "Opportunities and Challenges of Peer-to-Peer Internet Video Broadcast," *Proc of IEEE*, vol 96, pp 11-24, 2008.

[8] L. Posa, "Hamiltonian circuits in random graphs", *Discrete Math.*, vol. 14, pp. 359-364, 1976.

[9] N. Sarshar, V. Roychowdhury. "Scale-free and stable structures in complex *ad hoc* networks," *Physical Review E*, vol. 69, pp. 026101, 2004.

[10] E. Weisstein. "Grid Graph," From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/GridGraph.html