

Measurement of a Large-scale Overlay for Multimedia Streaming

Long Vu, Indranil Gupta, Jin Liang, Klara Nahrstedt
Department of Computer Science, University of Illinois at Urbana-Champaign, IL61801
{longvu2,indy,jinliang,klara}@cs.uiuc.edu

Categories and Subject Descriptors

C.2 [Computer-Communication Networks]

General Terms

Measurement, Performance

Keywords

Overlay Structure, Peer-to-peer media streaming, Modeling

1. INTRODUCTION

PPLive is currently the most well-known instance of an IPTV (Internet Protocol Television) application which stands out due to its increasing popularity. As of May 2006, PPLive had over 200 distinct online channels, a daily average of 400,000 aggregated users, and most of its channels had several thousands of users at their peaks. During the Chinese New Year 2006 event, a particular PPLive channel had over 200,000 simultaneous viewers [2].

There are several measurement studies about PPLive characteristics [1, 2, 5]. These existing studies tend to predominantly look at either network-centric metrics (e.g., video traffic, TCP connections, etc.), or at user-centric metrics (e.g., geographic distribution, user arrival and departure, user-perceived quality, channel population, etc.). Our crawler-based measurement studies are unique in focusing primarily on *overlay-based characteristics*, which lie somewhere in between the user-centric view and the network centric view. Our studies reveal that (1) node degree is independent of channel population size, and (2) small PPLive overlays (as many as 500 nodes) are similar to random graphs in structure.

2. PPLIVE BASICS

PPLive Protocols. When a PPLive node joins the network, it performs following steps: (1) retrieve a list of channels from channel management servers; (2) for the interested overlay, retrieve a small set of member nodes from the membership servers via *UDP*; (3) use this seed partner list to harvest (learn about) other candidate partners by periodically probing existing partners (and sometimes membership servers) via *UDP*. Further details are in [2, 3].

PPLive Overlay and Node Degree. We formally define the PPLive overlay as a graph $G = (V, E)$. Recall that each PPLive overlay corresponds to an individual PPLive channel. Each node (or peer) is defined as a given $\langle IP, port \rangle$ tuple and belongs to

Copyright is held by the author/owner(s). This work was in part supported by NSF CNS grant 05-09314.
HPDC'07, June 25–29, 2007, Monterey, California, USA.
ACM 978-1-59593-673-8/07/0006.

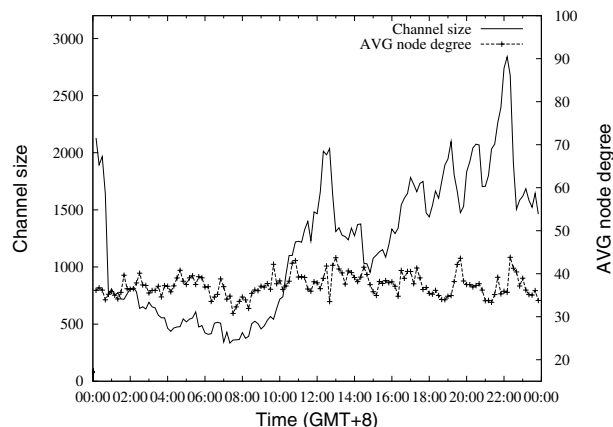


Figure 1: Average node degree is independent of channel population size (or channel size)

V. Each partner (or neighbor) of this node, appearing in its partner list, then corresponds to an edge (or link) in E .

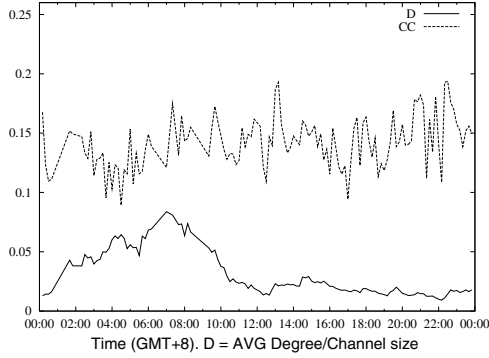
k response degree. We are interested only in the out-degree of a node in the overlay, and henceforth call this simply as “degree”. The *k response degree of a node* is defined as a aggregated set of partners in the first k responding packets from a node. In other words, when node gets queried repeatedly for its partners, it returns its partners and they are aggregated.

3. STUDY METHODOLOGY

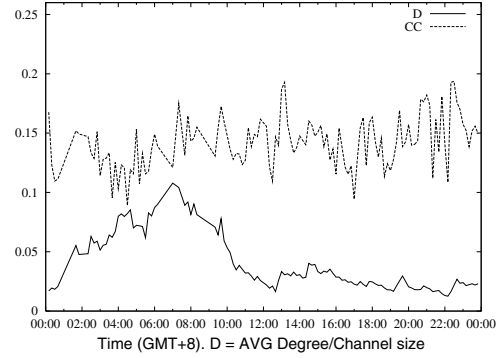
Settings. Our crawler works in the following manner: a machine (either a Linux machine in our CSIL cluster at UIUC, or a PlanetLab node) joins a given PPLive channel, and then crawls peers attending that channel. The crawler is used to execute what we call the *Snapshot Operation*. For the PlanetLab setting, we crawl using 10 geographically-distributed hosts, and aggregate their crawled information. Ethereal is used to trace traffic between PPLive nodes and servers.

Snapshot Operation. This operator attempts to capture the set of nodes present in the overlay over a short period of time ($O(\text{minutes})$). It works by repeatedly fetching partner lists and querying returned entries. Returned peers are aggregated.

Partner Discovery. This operation obtains the k response degree of a node. We repeatedly request a peer (once every second) to send its partner list. The first k received responses are aggregated to create the k response degree (and k response partner list).



(a) # of responses $k=5$



(b) # of responses $k=15$

Figure 2: Overlay resembles a random graph when channel size is small but becomes more clustered when channel size grows.

4. AVERAGE NODE DEGREE IS INDEPENDENT OF CHANNEL POPULATION SIZE

We collect the k response partner lists of nodes using our Snapshot Operation and Partner Discovery engine (Section 3). During a given snapshot operation (over 10 minutes), we simultaneously run the Partner Discovery to obtain the k response degree of 300 randomly selected peers that are both active and responsive. Figure 1 shows the variation of the average k response node degree and Channel Population Size (henceforth degree) of a channel during a 24 hour period, for $k = 15$.

These Figures first indicates that although the average node degree varies, it stays within a small range - between 30 to 43 over the course of day. More importantly though, *there appears to be no correlation between the variation of average degree and the channel size*. Thus we conclude that the average degree of a PPLive node does not depend on the channel population size.

This behavior can be explained since a peer in an overlay only needs to communicate once in a while with a few of peers to exchange data, advertise availability, and discover new partners. Therefore, even when the channel becomes large, a peer can still preserve viewing quality without establishing too many new connections.

5. RANDOMNESS OF OVERLAY DEPENDS ON CHANNEL POPULATION SIZE

The distinction between a random and a non-random graph can be measured by a metric called “Clustering Coefficient” (CC) [4]. Informally, the CC in a graph is defined as follows: for a random node u with two neighbors v and w in its partner list, the CC is the probability that either v is in w ’s partner list, or vice versa. The farther the CC is from the above value, the less random (i.e., more clustered) the graph is. For our experiment, we first calculate the average degree of the PPLive overlay (measured as in Section 4), and use it to calculate D as follows:

$$D = (\text{Average node degree}) / (\text{Channel size}) \quad (1)$$

We then compare the CC (measured as described below) to D (the unconditional probability that v links to w). The CC information is measured simultaneously to the degree measurement, and as follows: in each snapshot, we randomly select 300 active peers (i.e. root nodes). For each root node R , we first use partner discovery to obtain its partner list. Second, we pick randomly two active partners P_1 and P_2 in R ’s partner list and obtain their partner lists, again via the partner discovery operation. Third, we verify whether P_1 is

in P_2 ’s partner list or not, and vice versa. If P_1 is in P_2 ’s partner list (or vice versa), we increase a variable called *Count* by 1. *Count*, initialized to 1, represents the total number of edges existing in all such partner pairs. Then, CC is computed as follows:

$$CC = \text{Count} / (2 \times \text{RootNodeNum}) \quad (2)$$

In which, *RootNodeNum* is the number of root nodes whose two active partners P_1 and P_2 are verified. Figure 2 plots, for two different values of $k = 5, 15$, the 24-hour variation of D and CC . This experiment was done at the same time and for the same channel as Figure 1. Thus, notice that from 04:00 am to 08:00 am, when the channel population size is small, the value of CC approaches the value of D , especially for higher values of k . This indicates that *when Channel Population Size is small, the structure of the PPLive overlay graph approaches a random graph*.

This is explained by the fact that for “small” channels (with as many as 500 nodes), peers indeed connect fairly randomly to each other. As the channel population size increases (10:00 am onwards in Figure 2), the CC is only about six times that of the value of D . This is still indicative of some randomness of the graph, although it is clear large channel population sizes lead to more clustering.

6. CONCLUSION

We studied the overlay characteristics of the proprietary PPLive protocol, currently the world’s largest and most popular p2p media streaming system. Our studies reveal that (1) node degree is independent of channel population size, and (2) small PPLive overlays (as many as 500 nodes) are similar to random graphs in structure.

7. REFERENCES

- [1] A. Ali, A. Mathur, and H. Zhang. Measurement of commercial peer-to-peer live video streaming. In *Workshop in Recent Advances in Peer-to-Peer Streaming*, 2006.
- [2] X. Hei, C. Liang, J. Liang, Y. Liu, and K. W. Ross. A measurement study of a large-scale P2P IPTV system. Technical report, Polytechnic U., 2006.
- [3] L. Vu, I. Gupta, J. Liang, and K. Nahrstedt. Mapping the pplive network: Studying the impacts of media streaming on p2p overlays. Technical report, UIUC, 2006.
- [4] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 1998.
- [5] X.Hei, C.Liang, J. Liang, Y. Liu, and K. W. Ross. Insight into PPLive: Measurement study of a large scale P2P IPTV system. In *WWW 06 Wshp. IPTV Svcs. over WWW*, 2006.