CS 856 Paper Reviews
W Anthony Young - 20161423
October 26, 2004

Paper Title: Measurement, Modeling and Analysis of a Peer-to-Peer File-Sharing Workload

Section I. Overview

A. Reader Interest

1. Which category describes this paper?

(X) Practice / Application / Case Study / Experience Report
( ) Research / Technology
( ) Survey / Tutorial / How-To

2. How relevant is this paper to the readers in the area? Please explain your rating under III. Detailed Comments.

( ) Very Relevant
( ) Relevant
(X) Interesting - but not very relevant
( ) Irrelevant

B. Content

1. Please explain how this paper advances this field of research and / or contributes something new to the literature. Please explain your answer under III. Detailed Comments.

2. Is the paper technically sound? Please explain your answer under III. Detailed Comments.

(X) Yes
( ) Appears to be - but didn't check completely
( ) Partially
( ) No

C. Presentation

1. Are the title, abstract, and keywords appropriate? Please explain your answer under III. Detailed Comments.

(X) Yes
( ) No

2. Does the paper contain sufficient and appropriate references? Please explain your answer under III. Detailed Comments.

(X) References are sufficient and appropriate
( ) Important references are missing; more references are needed
( ) Number of references are excessive

3. Does the introduction state the objectives of the paper in terms that encourage the reader to read on? Please explain your answer under III. Detailed Comments.

(X) Yes
( ) Could be improved
( ) No

4. How would you rate the organization of the paper? Is it focused? Is the length appropriate for the topic? Please explain your answer under III. Detailed Comments.

( ) Satisfactory
(X) Could be improved
( ) Poor

5. Please rate and comment on the readability of this paper.
Please explain your answer under III. Detailed Comments.

(X) Easy to read
( ) Readable - but requires some effort to understand
( ) Difficult to read and understand
( ) Unreadable

Section II. Summary and evaluation

A. Summary (provide here a summary of the paper)

This paper presents an overview of the traffic workload of Kazaa, a popular file-sharing network on the internet. Kazaa experiences fetch-at-most-once retrieval semantics for files, where users download a file only once. This is in sharp contrast to WWW traffic.

The authors' analysis shows that Kazaa users are very patient when waiting for their downloads to finish: often waiting hours, days or weeks to retrieve files. Also, as Kazaa users age (i.e. use the program longer) they download less from the system.

Popularity of files tends to be short lived. Users typically download newer files than older ones. Kazaa download patterns also contradict a Zipf distribution. Thus, download patterns differ significantly from those of WWW documents. The authors proved through random simulation that fetch-at-most-once traffic does not follow Zipf distributions.

The authors propose that locality be employed in future versions of Kazaa in order to allow users to download from peers that are close (in terms of network distance) from them. Thus, a notion of caching could be used to reduce network strain and simultaneously boost performance.

B. Evaluation

Please rate the paper. Please explain your answer under III Detailed Comments.

( ) Award Quality
( ) Excellent
(X) Good
( ) Fair
( ) Poor

Section III. Detailed Comments

The authors present an abstract and keywords that describe the paper quite well. Their introduction provides an overview of the problem they are studying as well as an outline of the results to be presented. This paper contains complete and appropriate references to background material. However, missing is a discussion of the Kazaa routing protocol that could have aided in understanding some of the traffic patterns as well as some of the network behaviour.

This paper appears to be technically sound. The authors present their assumptions as well as collection and analysis methodology clearly. Assumptions appear to be valid. However, I do not believe that the information in this paper advances the field at all. It appears to be a characterization of traffic. As well, the main suggestion of this paper is to exploit locality of users to allow some notion of caching to be employed while downloading files. This is a notion that has been proposed before in other literature.

This paper is well organized and easy to read. However, the authors could have cut some material and repetitious comments. I believe that the paper could have been reduced in size while keeping the important information intact.

I also have a problem with the authors' characterization of the network load. Some points I wish to make are:

-Do users perhaps reduce the number of downloads because they initially download a lot of songs or videos that they really want, and once they have that data, they only download newer released songs and videos? There might be a correlation between download rate and age of data (i.e. initially, users download a lot of old files they really want for their "collection", and later, they download newer ones to add to it).

-Is it possible that download rates decreased due to changes in the amount of academic workload of students? Could you have a biased sample measuring only student traffic?

-Is it possible that users start downloading files from inside the network, which is not captured by your trace? That could be why your object popularity falls off? Not knowing the Kazaa routing protocol, I cannot speculate accurately if this is the case.

Overall, I believe this paper could have used some improvement.