

Enhancing Quality of Service by Exploiting Delay Tolerance in Multimedia Applications

Saraswathi Krithivasan, Sridhar Iyer
 Kanwal Rekhi School of Information Technology
 IIT Bombay, Mumbai, India
 saras,sri@it.iitb.ac.in

Categories and Subject Descriptors: C.2.3

General Terms

Algorithms, Performance, Design, Theory, Experimentation.

Keywords: Delay tolerant applications, Multimedia dissemination, Quality of Service (QoS), Transcoding, Caching, Heterogeneous networks, Distance Education.

1. INTRODUCTION AND MOTIVATION

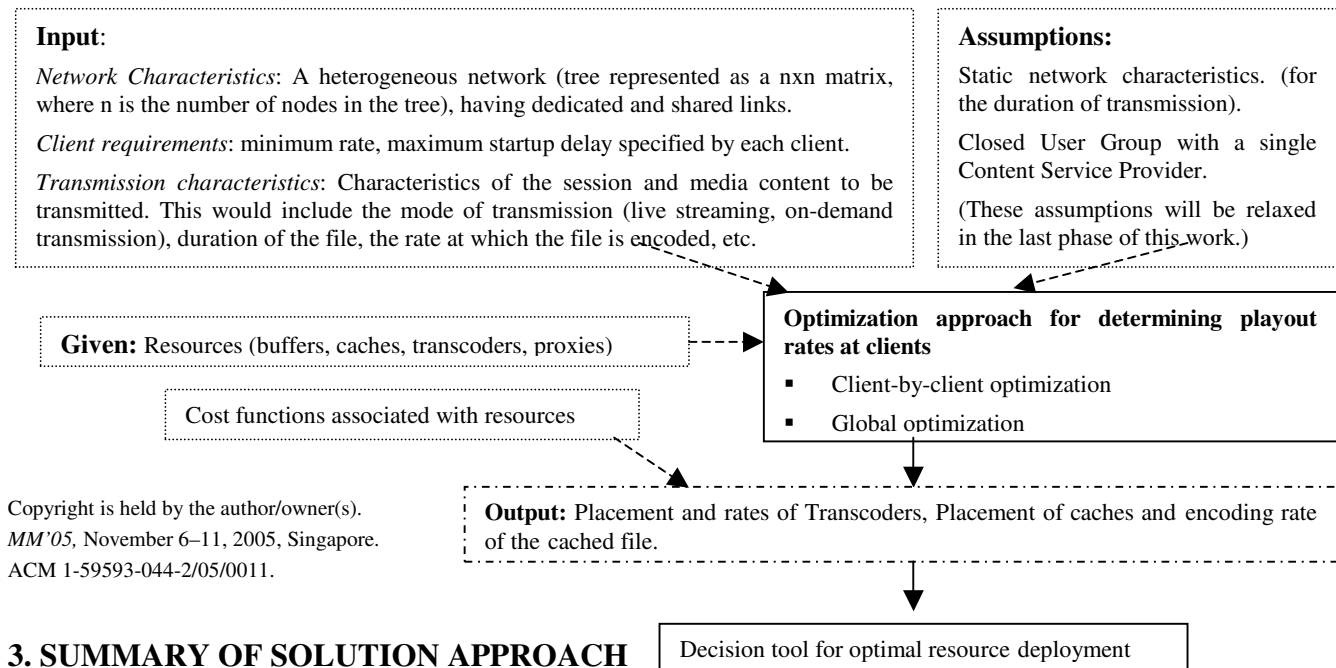
With the proliferation of world-wide computer networks, several popular streaming media applications have emerged: Universities offering their courses to a set of global subscribers, service providers streaming movies requested by their clients, and multinational corporations providing training to employees across cities. Heterogeneous architectures comprising of satellite, terrestrial links as well as the Internet are increasingly deployed for such applications. In these applications, a source disseminates multimedia contents that may be encoded at different rates to a set of geographically distributed clients through links of varying capacities and characteristics. In these applications the clients specify their QoS requirements in terms of a *minimum playout rate* they desire and a *maximum startup delay* they can tolerate. We define delay-tolerant applications as

those where clients are willing to wait for a specified time for the start of play back.

A review of the existing mechanisms for effective and efficient delivery of multimedia in [2][3] indicates that existing work treats multimedia dissemination as real-time applications that can tolerate some transmission errors and explores ways to *minimize* the startup delay. In contrast, we focus on multimedia applications that *can* tolerate delays.

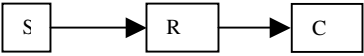
2. SCOPE OF WORK

In delay-tolerant applications, there is little advantage to be gained by serving the clients before their scheduled playout time. However, there are several benefits to the Content Service Provider (CSP) (due to enhanced customer satisfaction) in offering a playout rate that is better than what a client demands, if this can be achieved without jeopardizing the ability to satisfy the minimum client requirements and *without additional resources*. By maximizing QoS to the clients with given resources, CSPs can (1) maximize the utilization of links (2) provide differentiated services to their clients, and (3) offer upgraded services to some clients (which may have a revenue implication) without incurring any additional costs for resources. The scope of our work is summarized in the figure below:



Copyright is held by the author/owner(s).
 MM'05, November 6–11, 2005, Singapore.
 ACM 1-59593-044-2/05/0011.

3. SUMMARY OF SOLUTION APPROACH

Work completed: <i>Static network characteristics, Transcoding only</i>		
	<i>Description</i>	<i>Details</i>
1	<p>Basic block analysis:</p> <p>Analysis using a basic block consisting of a <i>source (S)</i>, <i>relay(R)</i>, and a <i>client (C)</i>, the three types of nodes used to represent any heterogeneous network.</p>  <pre> graph LR S[S] --> R[R] R --> C[C] </pre>	<p>The following three cases are considered:</p> <ul style="list-style-type: none"> • Startup delay= Infinity; The best playout rate r_{Base} can be provided to the client. • Startup delay= zero; Playout rate at client depends on the weakest link in its path. • Startup delay= d_i; Playout rate= At least r_i as defined by client C_i.
2	<p>Client-by- client optimization:</p> <p>Extension of basic block to path of client C_i, which specifies both r_i and d_i (Path from source to each client is considered in isolation).</p>	<p>Posed as an optimization problem, solved using Matlab's fmincon function; To handle large networks heuristics are also developed. Provides:</p> <ul style="list-style-type: none"> • Feasibility check for each client to ensure that both requirements can be met, given the network constraints. • Upper bound of the optimal rate that can be realized at client C_i.
3	<p>Global optimization:</p> <p>The entire network is considered. An additional constraint that accounts for multiple clients sharing links is added to the optimization problem.</p>	<ul style="list-style-type: none"> • Finds optimal rates for each client • Provides placement and rates of transcoders to realize the optimal rates.
Roadmap for remaining work: <i>Adding caching capability to nodes, catering to changing network characteristics.</i>		
4	Analysis with other adaptive mechanisms such as caches and combination of transcoders and caches.	<ul style="list-style-type: none"> • Caching at relay nodes and caching strategies to provide optimal rates. • Viability of combining mechanisms (transcoding and caching) at different sub-trees.
5	Study of trade-off between costs incurred for deployment of transcoders/caches and rates provided to the clients.	<ul style="list-style-type: none"> • Refine global optimization to include minimization of number of transcoders/caches and develop approaches to factor costs involved in deploying resources and the additional processing delays.
6	<p>Catering to:</p> <p>Links with dynamically varying capacities.</p> <p>Clients which dynamically join/leave the network.</p>	<ul style="list-style-type: none"> • Explore how our optimization approach can be tailored to work with QoS mechanisms that use adaptive techniques based on bandwidth availability [1][3] in dynamic networks. • Study the impact on the choice of mechanisms and optimal rates at the client when clients exhibit dynamic behaviour.
7	Implementing the tool: Includes the design, protocol between the server and client modules.	<ul style="list-style-type: none"> • Centralized design: Having a server module that takes decisions on the mechanisms to be invoked and a client module that takes parameters from the server module and invokes the appropriate mechanisms. • Distributed approach: Here decisions are made at different levels of the tree. Additionally, nodes where decisions are made (strategic nodes from where sub-trees emanate) need a protocol to coordinate their decisions.

4. CONCLUSION

With our preliminary simulation experiments, we have shown that clients that can tolerate startup delays can be served with higher rates with little or no additional resources. We have formulated our solution as an optimization problem to serve clients in a heterogeneous network with optimal rates. In addition to the development of the optimization tool, our proposed work will provide a decision tool for optimal deployment of resources to best serve multimedia content to a set of delay tolerant clients in a heterogeneous network.

5. REFERENCES

1. J. Liu, B. Li, Adaptive Video Multicast over the Internet, IEEE Multimedia, January-March 2003.
2. S. Krithivasan, Mechanisms for Effective and Efficient Dissemination of Multimedia, Technical report –September 2004, URL: www.it.iitb.ac.in/~saras
3. X. Wang, H. Schulzrinne, Comparison of Adaptive Internet Multimedia Applications, IEICE Transaction Communication, VOL.ES2-B, NO.6, June 1999.