Using Service Proxies for Content Provisioning

Panagiotis Kalliaras
National Technical University of Athens, Greece

Anthanasios-Dimitrios Sotiriou
National Technical University of Athens, Greece

INTRODUCTION

In modern broadband mesh networks, communication between two end nodes is carried out not directly, but through a number of intermediate nodes. While these nodes' only function may be to relay information from one point to another, they may also host computational elements which perform some service on behalf of other applications. We deal with the problem of optimally mapping multimedia content transcoding service elements onto network resources. There may be several places in the network where the required compression and decompression services could be performed. We would like to select the best locations that meet the application’s requirements. We propose a new approximation algorithm for constrained path optimization, which provides better scalability and simplicity than previous approaches. This is accomplished basically by partitioning the overall problem into smaller ones.

RELATED WORK

The majority of the proposed schemes are focused on solving the similar multi-constrained optimal-path problem (MCOP). This problem aims to find in a network an optimal path that satisfies multiple additive path constraints and has been proven to be of NP-complete complexity, therefore unsolvable in polynomial time. Several algorithms have been proposed for the above problem. For the MCOP problem with two parameters, Jaffe (1984) proposed to use a linear weight combination of the two constraint parameters. Other proposed algorithms include Iwata et al. (1996), SAMCRA (Van Mieghem, De Neve, & Kuipers, 2001), and the Chen-Nahrstedt algorithm (Chen & Nahrstedt, 1998). All of the above algorithms require a global state to be maintained at every node. Most algorithms transform the routing problem to a shortest path problem and then solve it by Dijkstra’s or the Bellman-Ford algorithm.

The concept of service path has been proposed in smaller numbers. In TranSquid (Maheshwari, Sharma, Ramamirtham, & Shenoy, 2002), a transcoding and caching proxy for heterogeneous clients is proposed. In the Ninja Project (Gribble et al., 2001), service path is defined as a sequence of application-level service operators and connectors. In Lienhart, Holliman, Chen, and Yeung (2002), the authors propose the addition of a media support module (MAPS) on top of an existing peer-to-peer service layer, in order to improve multimedia services across heterogeneous computing platforms. This module is responsible for transcoding and route path selection based on the single-pair shortest-path problem and utilizes Dijkstra’s algorithm to provide a solution. Our system is a combination of the above two research areas. It uses the service paths concept and also fully implements the MCOP, rather than the simpler path-finding solutions.

NETWORK AND SERVICE MODELING

We assume that our network topology matches that of a partial mesh network. A mesh network is reliable and offers redundancy. If one node can no longer operate, all the rest can still communicate with each other, directly or through one or more intermediate nodes. Mesh networks work well when nodes are located at scattered points that do not lie near a common line.

As a service, we denote any network resource which may include computational elements and performs some online activity on behalf of other applications. A service is an online facility that is always available to all requestors, at a predefined cost and delay. The services may be available on more than one node, either serially or concurrently. Although services in communication networks delivering multimedia content may include conversion processes like media data adaptation, merging of multiple media sources, copyright protection, metadata extraction, enhancements, and recovery, in our modeling we focus mainly on transcoding. Possible forms of transcoding include: lowering the bit rate of a media stream by reducing the image/video resolution, size, and/or frame rate; converting a media stream from one encoding format to another; or a combination of the above. The expected benefits from these adaptations are on one hand to move computation (data transformation) from the client site to the proxy, and on the other hand to reduce the volume of data transferred to the client. In any case, a service in our modeling accepts a media stream which is...
Using Service Proxies for Content Provisioning

characterized from an input data rate, performs application-
level processing on the media data, and forwards the stream
at a new output data rate.

The connection requests are initiated by clients that
wish to acquire specific content from content providers. The
procedure that the clients follow to detect content provid-
ers with desirable content is out of the scope of this article.
The notion of clients, content providers (CP), and service
providers (SP) is used here to state the dynamic nature of
the network connections. The scope of multimedia service
provision is to provide clients with customized and satis-
factory QoS, under the constraint of end-to-end resource
availability observed by each client. Several requests are
initiated, targeting various service entities.

The main pattern for the proposed solution is to split
the problem into smaller ones, as shown in Figure 1. Each
request forms two bids. The first bid concerns the path from
the request to the service that has the desired output. This
bid includes the network and service paths, and the delay
and cost for the path and the usage of the service. The sec-
ond bid is required if the chosen service does not own the
content, and acquisition is needed from somewhere else. In
this case, a new request is formed and its result is returned
to the initial request. The results from the second bid con-
tain everything after the service, including any subsequent
services that may be used.

PROBLEM FORMULATION
AND HEURISTICS

Let directed, connected graph \( G(V, E) \) denote the network
topology, where \( V \) is the set of vertices of the graph (repre-
senting network components (e.g., switches, routers, hosts,
aggregated subgraphs) and \( E \) the edges (representing com-
munication links). There are four basic entities used in our
problem formulation, shown in Table I. Edges and vertices
are part of the actual network and services, and requests are
part of the overlay network.

PROBLEM ENTITIES

<table>
<thead>
<tr>
<th>Edge ( e )</th>
<th>Vertex ( v )</th>
<th>Service ( s )</th>
<th>Request ( req )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B_e ): available bandwidth</td>
<td>( d_v ): delay</td>
<td>( v_r ): host vertex</td>
<td>( v_{req} ): host vertex</td>
</tr>
<tr>
<td>( d_s ): delay</td>
<td>( r_{in}/r_{out} ): input/output rate</td>
<td>( D_{max} )</td>
<td>max delay</td>
</tr>
<tr>
<td>( c_s ): cost per rate</td>
<td>( c_r ): cost</td>
<td>( R_{req} ): requested rate</td>
<td></td>
</tr>
<tr>
<td>( d_e ): delay</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For every request, ensure the end-to-end delay constraint
is met:

\[
\sum_{e \in P} d_e + \sum_{e \in SP} d_s + \sum_{e \in SP} d_e \leq D_{max}^{req}
\]  

(1)

while the available bandwidth for all edges in the path is at
least the required rate at that point:

\[
R_{req} \leq B_e, \forall e \in P
\]

(2)

and minimize cost:

\[
C_{req} = \sum_{e \in P} C_e \cdot r_e + \sum_{e \in SP} c_e
\]

(3)

where P is the network path and SP the service path.

The problem, as described above, is more accurately
described as unicast link-constrained path-constrained
path-optimization routing. The link constraint refers to the
available bandwidth, the path constraint to the total delay,
and the desirable path optimization to the minimization of
the total cost.

The Heuristics Solution

The main steps for estimating the optimal paths are the
following:

- **Step 1. Topology Filtering:** For every request \( req \)
  initiated at host vertex \( t \) and inquiring an object with
  optimal rate \( R_{req} \) and maximum delay \( D_{max} \)
  filter all links (and possibly disconnected nodes) that do not
  satisfy the requested minimum linear QoS constraints,
in our case, minimum available bandwidth \( B_e \) as in
  equation (2).
- **Step 2. Finding Available Services:** Let \( S_s \in S \) be the
  sum of services that have an output rate \( R_{req} \)—that is,
Related Content

Projected Displays of Mobile Devices for Collaboration
[www.igi-global.com/chapter/projected-displays-mobile-devices-collaboration/21854?camid=4v1a](www.igi-global.com/chapter/projected-displays-mobile-devices-collaboration/21854?camid=4v1a)

Development of an Information Quality Framework for Mechanical Engineering Modules with Enhanced Treatment for Pedagogical Content
[www.igi-global.com/article/development-of-an-information-quality-framework-for-mechanical-engineering-modules-with-enhanced-treatment-for-pedagogical-content/175345?camid=4v1a](www.igi-global.com/article/development-of-an-information-quality-framework-for-mechanical-engineering-modules-with-enhanced-treatment-for-pedagogical-content/175345?camid=4v1a)

Automatic Signature Verification on Handheld Devices
[www.igi-global.com/chapter/automatic-signature-verification-handheld-devices/38547?camid=4v1a](www.igi-global.com/chapter/automatic-signature-verification-handheld-devices/38547?camid=4v1a)

Visualizing Pathway on 3D Maps for an Interactive User Navigation in Mobile Devices
[www.igi-global.com/chapter/visualizing-pathway-on-3d-maps-for-an-interactive-user-navigation-in-mobile-devices/208463?camid=4v1a](www.igi-global.com/chapter/visualizing-pathway-on-3d-maps-for-an-interactive-user-navigation-in-mobile-devices/208463?camid=4v1a)