A Market-Based Approach to Allocating QoS for Multimedia Applications

Hirofumi Yamaki
Department of Information Science
Kyoto University
Kyoto, 606-01, Japan
yamaki@kuis.kyoto-u.ac.jp

Michael P. Wellman
Artificial Intelligence Laboratory
University of Michigan
Ann Arbor, MI 48109-2110 USA
wellman@umich.edu

Toru Ishida
Department of Information Science
Kyoto University
Kyoto, 606-01, Japan
ishida@kuis.kyoto-u.ac.jp

Abstract

Allocating quality of service (QoS) has been a focus of recent work on distributed multimedia systems and networks. This paper explores a decentralized approach that allocates QoS through a dynamic market. In our approach, each agent makes decisions according to its local knowledge and interests, and prices adjust to clear the market in each resource. Dynamic changes in agent needs and network status cause the agents to revise their decisions continually. The market prices reflect system-wide values, inducing agents to produce and consume appropriate amounts of the various resources. We describe a market model for allocating bandwidth in an actual networked meeting environment called FreeWalk. Experiments reveal the responsiveness of the market-based approach to dynamic conditions in allocating QoS for multimedia network applications.

Introduction

Modern multimedia applications can make use of nearly unbounded amounts of network bandwidth. Allocating more bandwidth to the application can produce improvements in latency, fidelity, resolution, reliability, or other important service features. Regardless of the bandwidth available, there is almost always a potential value to having more devoted to the application.

Since the available bandwidth may be distributed across the possible applications and uses in a variety of ways, there are always tradeoffs in allocation. The allocation policy implemented by the network determines the quality of service (QoS) provided to each task in the distributed system. This problem of QoS allocation in a distributed environment has been an area of focus in recent work on multimedia systems and networks (Nahrstedt95; Vogel95).

Although our ideal allocation would produce an optimal overall quality of service, our mechanism for computing this allocation must account for the distribution of relevant information (e.g., network loads, application characteristics, quality preference tradeoffs, time-phased service demands) and computing power, as well as the decentralization of decision making authority.

One approach that focuses particularly on this decentralization issue is to allocate resources through a market. In a market-based mechanism (Clearwater95), the participating decision makers, or agents, exchange resources at established market prices. Each agent makes decisions according to its local knowledge and interests, and prices adjust to clear the market in each resource. Under certain conditions, the market prices reflect system-wide values, inducing agents to produce and consume appropriate amounts of the various resources.

In this research, we applied a market-based approach for allocating bandwidth in a networked multimedia meeting environment called FreeWalk (Nakanishi96). FreeWalk is the product of a broader project, called Socia (Ishida94; Yamaki96), aimed at supporting human communities through computer networks.

Since QoS requirements change dynamically in FreeWalk, our market model focuses on the intertemporal dimension of the allocation. Our simulation results show that (1) the market-based approach enables each FreeWalk client to respond appropriately to dynamic changes in network loads and client locations, and (2) agents with different current and future needs for bandwidth can effectively trade across time periods to achieve an allocation that makes them all better off.

FreeWalk Meeting Environment

Meetings in FreeWalk

With the advance of computer networks have appeared numerous systems supporting collaborative work. Perhaps the most common type of tool are desktop conferencing facilities designed to support formal business meetings. However, meetings are not always for business, nor are they always formal. Casual meetings such as chatting at a coffee break or in a passageway enrich our life, and play an important role in collaboration.
Casual meetings are characterized by accidental encounters, unlimited participants, and unpredictable topics of conversation. Conventional desktop conferencing tools are not ideal for these kinds of meetings. For example, displaying the faces of all participants can strain conversation, and limits the number of people that can join at once. FreeWalk, in contrast, is designed to support an informal atmosphere—like a park or lobby—where people meet through accidental encounters or purposeful gatherings. By providing for self-directed grouping within a common virtual space, larger numbers of participants can be accommodated.

People enter the FreeWalk space by connecting to the server. In this virtual community common, each participant is represented as a pyramid of 3-D polygons. A live video image is mapped on one rectangular plane of the pyramid. The participant’s viewpoint is located at the center of this rectangle. The view of the community common from this viewpoint is displayed in the FreeWalk window, as shown in Figure 1. Participants standing far away appear small, and those nearer are larger. Participants located outside of a predefined distance are not displayed. Similarly, the volume of voice is proportional to the distance between sender and recipient. Participants navigate around the FreeWalk plane by driving their image around using a mouse—just as in a video game. People can find the locations of other participants using the radar screen in the lower-right corner of the window.

In FreeWalk, people can show up in a meeting space, wander freely inside the space, and encounter each other accidentally. Since the locations and view directions of the participants are reflected by pyramid orientation, each can watch what the others are doing from a distance. Since people can grasp what is going on in the community common at first glance, many participants can simultaneously exist in the same space without confusion. This feature makes FreeWalk an effective tool for holding a party with more than five people.

**FreeWalk QoS Problem**

The FreeWalk system consists of a community server and clients, each of which manages vision and voice processes. Figure 2 illustrates the interaction between the community server and clients. When a participant makes a move by using a mouse, the corresponding client calculates the new location and orientation, and sends this data to the community server. The server compiles the global map and transmits it to each client for screen updating. Since only control information is transferred between the server and clients, the community server can efficiently maintain a global view of the ongoing activities in the community common.

Since the spatial relations of clients continuously change in FreeWalk, it is not effective to multicast the same pictures and voices to all other clients. Instead, FreeWalk clients send them in a more targeted fashion. Specifically, each client uses the map information to determine which participants have it in view, and sends the image only to those clients. Furthermore, the client adjusts the shape of its owner’s picture to the texture-mapped plane in the receiver’s display. It then sends these adjusted pictures, controlling transmission rate based on the drawing speed of the receiving client.

Similarly, when the voice process receives all other clients’ addresses and locations, it sends the owner’s voice to other clients within the limit of hearing. When receiving voices from other clients, the process deter-
mines the volume of each voice based on the distance between them, and then composes all voices from other clients.

In the current FreeWalk implementation, though each client can control the amount of data transfer for each frame, the overall bandwidth for transferring those data is determined by the local network environment: traffic congestion uniformly restrains data transfer among clients. Therefore, as the number of clients increases, the network QoS experienced by FreeWalk clients can degrade significantly.

The major problem is that because bandwidth allocation across clients is fixed, the system cannot respond appropriately to dynamic changes in network loads and client locations. For example, there is no mechanism to allocate more bandwidth to clients that are interacting closely, as opposed to those that are wandering alone in FreeWalk space. As a result, QoS can vary suddenly during an interaction, and there is nothing a client to do about it. We believe that some mechanism for adaptive QoS control is therefore required for robust support of multi-user FreeWalk sessions.

Market Model for Bandwidth Allocation

Our adaptive QoS control mechanism is based on a market model. In this model, FreeWalk clients are agents in the system, bidding for bandwidth allocation according to their needs and means. Our aim is to provide a principled mechanism for bandwidth to be allocated toward those clients with the highest priorities and most effective uses for the bandwidth, responsive to the dynamic changes of clients distributed across the network. This model has been implemented in WALRAS, a market-oriented programming environment providing facilities for specifying and running computational economies (Wellman93).

Basic Concepts for Market Configuration

Figure 3 depicts our market model for bandwidth allocation in FreeWalk. Rectangles in the middle represent goods exchanged inside the market. The network goods are divided into bandwidth and QoS, and there are two different time periods, current and future. This means we have four types of goods in the market. Circles in the diagram represent agents, of two types. Consumer agents represent FreeWalk clients, and producers represent the current and future network operations. Directed edges indicate the flow of goods in the economy.

In Figure 3, \( BW \) and \( FBW \) denote current and future bandwidth, respectively. Current QoS on the connection from client \( j \) to client \( i \) is denoted \( q_{ij} \). \( F_q \) denotes future QoS available to client \( i \). Below, we describe two basic concepts underlying our market model.

1. **Clients primarily value FreeWalk QoS, rather than raw bandwidth.**

   Although bandwidth may be a reasonable proxy for the quality of FreeWalk service obtained by a client, we can describe the service level more directly in terms of QoS parameters. The QoS model recognizes that bandwidth is really a measure of the raw network resource, and what a client really cares about at the application level is how well this network supports the task at hand. Depending on the application, QoS can be represented in a variety of ways, for example video resolution, frame rate, sound quality, and so on.

   Thus, we treat bandwidth and QoS as different goods, and further distinguish QoS on each individual FreeWalk connection. The QoS goods are "produced" from bandwidth by system agents—QoS producers—as described in the next section.

2. **By distinguishing the "current" and "future" networks, we provide incentives for the relatively inactive clients to transfer network resources to their more active counterparts.**

   If the FreeWalk clients are interested in only the current network situation, they will simply choose how much of their allocated bandwidth to use for their respective FreeWalk connections. Thus, no actual
trading of bandwidth among the consumers takes place, even though each consumer's decision affects the others. Without an incentive to trade (or a power to force transfers), there is no opportunity to take advantage of the fact that bandwidth allocated to some agents may have more social value than bandwidth allocated to others.

A central feature of the FreeWalk system is that the agents move around the FreeWalk space in real time. As the agents move, their needs for bandwidth fluctuate. For example, when engaged in a meeting with one or more other clients the value of bandwidth is high, but while inactive or in transit the demand is low. An agent that could take a long-range perspective would attempt to garner more "future" resources for the active times, and to do so would be willing to give up "current" network resources, while another agent sells future resources to buy current resources.

By including current and future goods in the same price system, we enable the agents to effectively trade across time periods, according to their configuration inside the FreeWalk community space.

**FreeWalk Clients and QoS Producers**

Consumer agents in the market model correspond to FreeWalk clients. Client i's relative preference for bandwidth and QoS is represented by a utility function. For convenience, we adopt the CES (constant elasticity of substitution) preference model. As above, we represent QoS on the connection from client j to client i as \( q_{ij} \). Denoting the amounts of future bandwidth, future QoS, and current bandwidth allocated to client i by \( x_{i}^{bw}, x_{i}^{qos}, \) and \( x_{i}^{cbw} \) respectively, the CES utility function of client i as a consumer is given by

\[
\begin{align*}
\bar{u}_i(x_{i}^{bw}, x_{i}^{qos}, x_{i}^{cbw}, q_{i1}, \ldots, q_{im}) &= \left( \frac{\alpha_i x_{i}^{bw} x_{i}^{qos}}{\alpha_i x_{i}^{qos} x_{i}^{bw}} \right)^{\frac{\sigma-1}{\sigma}} \\
&\quad + \alpha_i x_{i}^{cbw} x_{i}^{qos} x_{i}^{bw} + \sum_{j=1}^{m} \alpha_{ij} q_{ij} x_{i}^{bw} \right)^{\frac{\sigma-1}{\sigma}} (1)
\end{align*}
\]

In the CES functional form, the \( \alpha \) coefficients dictate the relative values, and the global substitution parameter \( \sigma \) specifies the degree to which consumption in one good (at proportions dictated by the \( \alpha \)'s) can substitute for the others.

A consumer's ability to satisfy its utility depends on its endowment: the initial allocations of each good. In general, a consumer with endowment \( e = (e_1, \ldots, e_K) \) of the \( K \) goods chooses the consumption vector \( x = (x_1, \ldots, x_K) \) that solves the following optimization problem at prices \( p = (p_1, \ldots, p_K) \).

\[
\max_{x} u(x) \quad \text{subject to} \quad p \cdot x \leq p \cdot e.
\]

In our model, the endowment of a consumer includes no QoS; all service quality must be created by QoS producers, which are shown on right side of Figure 3. The input to the production process is bandwidth (although it could include other network resources such as buffer size or switching capacity), and the output is generic QoS. The role of the QoS producer in this economy is to describe the relationship between network resources and service quality, however it may be measured.

---

1CES forms are commonly employed in general equilibrium modeling (Shoven92), due to their flexibility and convenient analytical properties. The assumption underlying CES is that the ratio of fractional increase in one good's consumption that would compensate for a fractional decrease in the other is a constant independent of the consumption levels.
The relationship between bandwidth and QoS can be defined in various ways. The specific model we adopt describes current QoS $q_{ij}$ as the output of a production function,

$$f_{ij}(x) = Q_i - \frac{Q_i}{1 + \gamma_{ij} x},$$

(2)

where $Q_i$ is the maximum QoS that can be achieved by the FreeWalk process corresponding to client $i$. The parameter $\gamma_{ij}$ is proportional to the distance between clients $i$ and $j$, which makes the QoS of the corresponding connection change slowly against the allocated bandwidth when they are far from each other and faster when they come near. The production function for future QoS, $f$ (no subscript), is the same as (2), with its $\gamma$ fixed at unity.

The specific form of the function is not important, as long as the relationship exhibits decreasing returns, and has sufficient parameters to capture dynamic behavior in the FreeWalk environment. This is quite realistic in the usual operating range—after a point increasing the network resource yields diminishing amount of detectable improvement in the service quality.

Producers select their activity level to maximize profits,

$$\max p \cdot y \text{ subject to } y \in Y,$$

where $Y$ denotes the producer's technology, or set of feasible production plans. For current QoS producers, $Y$ consists of pairs $(x_{cbw}, q_{ij})$ such that $x_{cbw}$ is sufficient bandwidth to produce $q_{ij}$ units of QoS. For the specific model above, this means $q_{ij} \leq f_{ij}(x_{cbw})$.

For this class of economies, the WALRAS distributed bidding protocol (Cheng96) is guaranteed to converge to the unique competitive equilibrium. By setting the utility and production coefficients, we can calibrate the model to a baseline allocation we consider reasonable. For any such settings, the unique competitive equilibrium is Pareto optimal, and we can achieve any Pareto optimum by selecting appropriate endowments.

**Dynamic Resource Allocation**

To define fully the relationship between the two time periods, we must specify how to convert future network resources into current resources as time passes. We employ a rolling horizon approach to run our dynamic market model in WALRAS.

The model of the future network is based on an aggregation of future time slices (of some fixed duration), over a specified time horizon. Let $T$ denote the time horizon. Then the first time slice is represented by the current market period, and the future market period represents the remaining $T - 1$ slices. Given an overall network size of $\beta$, we have a total endowment of $\beta$ units of bandwidth in the current period, and $(T-1)\beta$ units in the future period. We must then set the preference and production parameters for the future network to calibrate them to this larger scale.

We run the model in WALRAS, and use the resulting values for current bandwidth and QoS as our network allocation. Then we increment the "clock" one time slice and repeat the process. In rolling time forward, however, we must account for the agents’ exchanges of future resources (otherwise, the future markets would be illusory, and the agents would not be behaving in their real interest). To do so, we determine each consumer’s share of the future resources, and use this to set the endowments of current and future resources for the next iteration of the model.

To determine consumer $i$’s share of future resources, $r_i$, we sum the equilibrium consumption of future bandwidth, $x_{i}^{bw}$, and the bandwidth that $i$ is effectively deploying to produce its consumption of future QoS. This latter quantity is derived by taking the total future bandwidth used by the future producer, and dividing it in proportion to the share of future QoS consumed. The overall result is given by

$$r_i = x_i^{ bw} + f^{-1}(x_i^{ f w}).$$

In equilibrium, $\sum_i r_i = (T - 1)\beta$, the total bandwidth available in the future. If we stop the bidding process before equilibrium is reached (the normal case), the $r^i$ are normalized to satisfy this material constraint.

---

2 To verify this, we have also experimented with a quadratic-cost technology, where the amount of bandwidth required to produce a particular QoS level is quadratic in that level. The results are qualitatively similar. As for our utility model, this production model is chosen primarily for analytical convenience rather than based on any empirical or theoretical analysis of the FreeWalk system.

3 If we stop the bidding process before equilibrium is reached (the normal case), the $r^i$ are normalized to satisfy this material constraint.
the clients move and the preferences change suggest that total utility (sum over agents over time) increases when agents have the opportunity to trade bandwidth across time periods.

**Experimental Evaluation**

**Scenario**

To test the market model, we simulated a scenario where a moving client resides far from other two fixed clients at first, and then moves toward them. The initial configuration of the experiment is shown in Figure 4.

Since the moving client does not need network service when it is far from the others, the fixed clients use all the bandwidth to generate QoS between themselves. As the moving client gets nearer, it starts to require more network service. In this experiment, therefore, the moving client is expected to buy future bandwidth and future QoS at first, and begin to sell them and buy current bandwidth and QoS as it approaches the fixed clients.

To run the scenario, we implemented a simulator that can evaluate various situation in FreeWalk, including that described above. In the simulator, we specify the location and orientation of the fixed clients within the FreeWalk plane, and the trajectory (location and orientation as a function of time, t) of the moving client. Given these parameters and the initial time, the simulator calculates the relative location of each client and determines which other clients are within its sight range. On each cycle, the simulator sets the preference coefficients, endowments, and other parameters based on the relative locations and clients in view, and then runs an iteration of the FreeWalk economy in WALTAS.

The maximal current QoS of each client $Q_i$ is set to one, time horizon $T$ to ten periods, and the overall network bandwidth $\beta$ to 30 bandwidth units. The CES utility coefficients are set to the following values:

$$\alpha_i^{bw} = \frac{1}{10},$$

$$\alpha_i^{f,bw} = \alpha_i^{f,qos} = \frac{\alpha_i^{bw}}{T},$$

$$\alpha_{ij} = \begin{cases} (s_{ij}s_{i0}) \times 100 & \text{if } j \text{ is inside the view of } i \\ 0 & \text{otherwise}, \end{cases}$$

where $s_{ij}$ is the size of client $j$'s live video plane in the FreeWalk window of $i$, and $s_{i0}$ is the total size of $i$'s FreeWalk window.

Each QoS producer produces current or future QoS using the technology defined by (2), with $\gamma_{ij}$ set to the distance between clients $i$ and $j$. The orientation of each client in the simulator is shown in the following table:

<table>
<thead>
<tr>
<th></th>
<th>Initial Location</th>
<th>Velocity</th>
<th>Angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed client 1</td>
<td>(-1, 0)</td>
<td>(0, 0)</td>
<td>$\pi/4$</td>
</tr>
<tr>
<td>Fixed client 2</td>
<td>(1, 0)</td>
<td>(0, 0)</td>
<td>$3\pi/4$</td>
</tr>
<tr>
<td>Moving client</td>
<td>(0, 30)</td>
<td>(0, -1)</td>
<td>$-\pi/2$</td>
</tr>
</tbody>
</table>

**Results**

Figure 5 depicts the result of the simulation. The horizontal axis of each chart corresponds to time steps in the simulator. Figures 5(a) and 5(b) plot the satisfied demand of current and future QoS, respectively. Figure 5(c) plots the current bandwidth devoted to producing FreeWalk QoS, and Figure 5(d) the future bandwidth endowments at the start of each simulation cycle.

In these charts, solid and dotted lines are of the moving client and the fixed clients respectively. The data for the fixed clients are merged into one line, because the orientations of these clients are symmetric and thus there is no difference between them.

The basic results obtained from our experiments agreed qualitatively with our expectations.

- The market allocation responds appropriately to dynamic changes in network loads and client locations.

---

4The unit length in the community common of the simulator is defined as the half of the width of a FreeWalk client's live video plane.
The simulation results show that the moving client buys current bandwidth by selling future bandwidth as it approaches the fixed clients (Figures 5(a) and 5(c)). Since the utility coefficient $a_{ij}$ is proportional to the displayed size, relative preference for QoS increases as the clients get nearer. The result is a greater tendency to trade bandwidth for QoS, and future for current goods, both of which are appropriate in the FreeWalk context.

- **Distinguishing the current and future networks enables intertemporal resource transfer among clients.**

When the moving client approaches to the other clients, endowed future bandwidth and future QoS is transferred from the moving client to the fixed client (Figures 5(b) and 5(d)). Though the simple two-period model is extremely coarse, it is sufficient to improve utility for all clients compared to the static model.

Scenarios quantitatively different from the instance described above yield qualitatively similar results. The magnitudes of dynamic fluctuations and intertemporal transfer depend on the trajectories of agents and the overall scarcity of the bandwidth resource.

### Related Work

The FreeWalk market model is the latest in a series of computational economies developed using the WALRAS market-oriented programming environment (Wellman93), and one of a growing number of applications of market-based approaches to distributed resource allocation problems (Clearwater96). Of the computational markets reported in the literature, several have been specifically devoted to allocating network bandwidth or other computational resources. For example, Kuwabara et al. (1996) present simulation results from a market-like model of communication network control. One of the more substantial efforts is that of Agorics, Inc. (Miller96), who have been developing auction infrastructure to support bandwidth allocation for a video-on-demand application. Several other projects focus on allocation of processing resources (Bogan94; Waldspurger92).

Our main innovation in the FreeWalk economy with respect to these prior works is the explicit introduction of futures markets in bandwidth and QoS. Most previous market-based systems allocate resources over time through repeated execution of spot markets for current resources. There are also some instances where
resources are reserved for particular future time slots including a very early auction-based scheme for processing time described by Sutherland (1968). The two-period rolling horizon method adopted for FreeWalk is intermediate between spot and reservation markets, and seems to offer an advantageous tradeoff of their respective desirable and undesirable features.

Conclusion

Networked multimedia applications like FreeWalk require efficient, dynamic, and decentralized techniques for allocating network resources. Our experience with the FreeWalk economy suggests that the market-based approach can support flexible QoS allocation in highly dynamic environments. To achieve these results, we expressly designed the model to reflect intertemporal agent preferences, and to be responsive to dynamic application conditions.

However, much remains to be learned about the performance and design of these techniques. Our ongoing and future work focuses on three areas:

1. Implementation of the market within the actual FreeWalk system (currently underway). Given on the existing communication channels used to transmit position information to the server, we believe that the bidding process will impose minimal additional overhead. Nevertheless, as we have found in our other deployment efforts, building the actual system always exposes unanticipated issues. Those we can anticipate include dealing with disequilibrium transactions, and gracefully accommodating client crashes and network delays.

2. Deeper models of service quality. The current model of QoS embodies an extremely simplified view of user preferences and network behavior. Whereas these simplifications may have been reasonable for validating the approach, incorporating markets in the actual FreeWalk system will require a higher level of realism.

3. Analysis of intertemporal allocations. We aim to construct principled design rules for determining how many futures markets to open, and for what time horizons, as a function of the FreeWalk configuration and expected dynamics. This will require a better understanding of the inherent tradeoff between fidelity and complexity in intertemporal resource allocation.

Acknowledgment We are grateful to the anonymous referees for suggestions about discussing related work, and other helpful comments.

References


