Chapter 6

A Spot Market Approach to Multi-User Allocation

As described in chapter 4, there are several advantages associated with microeconomic based, multi-user allocation. Many economic-based techniques offer a distributed allocation method, eliminating the need for a central controlling entity. Economic-based methods can also achieve optimal allocations (defined in section 5.2). These methods can scale to large networks and provide a framework for economic goals (such as, cost recovery and profit maximization). Furthermore, pricing network resources provides a disincentive to over-allocate resources. However, none of the microeconomic methods described in chapter 4 are able to handle network dynamics (changing user demands and users entering/exiting the network). When demands change, typically new prices must be calculated off-line. For this reason, these microeconomic-based methods are not suitable for VBR traffic sources.

In this chapter a multi-user allocation method based on the competitive market model is described. Referred to as a "spot market approach," the economic model consists of multiple *dynamic* competitive markets [38] (spot markets) working asynchronously and independently. Users are considered consumers and purchase link bandwidth to maximize their QoS. Switches are producers and sell link bandwidth at the current market price. This allocation approach has all the advantages associated with other microeconomic-based methods; however, the spot market approach allows and encourages network dynamics to

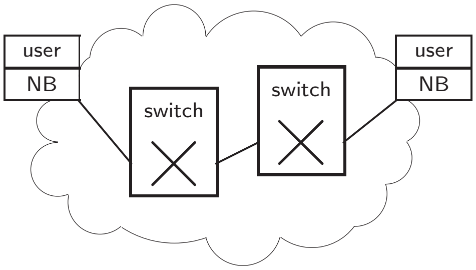


Figure 6.1: Example network consisting of users, network brokers and switches.

occur. In addition, the spot market moves the allocation calculations to the edge of the network. This reduces the computation and storage overhead required at the switches.

A detailed description of the spot market approach is provided in this chapter, followed by discussions concerning network dynamics and optimality. Simulation results will indicate that the spot market approach is able to achieve optimal and equitable allocations under network dynamics, performing better than other bandwidth allocation methods.

6.1 Spot Market Economy

This proposed network allocation method is based on a modified competitive market model (dynamic competitive market [38] or spot market), where pricing is done to promote high utilization of resources and Pareto-optimal allocations. The spot market has the unique ability to adjust to changing resource demands and provides users with immediate availability of resources.

As seen in figure 6.1, the network economy consists of three entities: users (those who execute network applications), Network Brokers (NB) and switches. Using the competitive market nomenclature, users are consumers, switches are producers and network brokers are used to assist the exchange of resources in the market. As mentioned in the introduction, there are many different types of network resources desired by users. This chapter will only consider link bandwidth; however, the techniques presented can be applied to other network resources.

6.1.1 Switch

In the economy, the switch owns the link bandwidth that is sought by consumers. The network consists of several switches interconnected with links. For a unidirectional link between two switches, consider the sending switch as owner of the bandwidth of that link. Each switch prices its link bandwidth based on local supply and demand for that link. Therefore a single switch, having multiple output links, will have one price associated with each output port. Bandwidth available in the spot markets is considered a non-storable resource (similar to residential electricity) and is available for immediate use (no reservation overhead required). The entire network can be viewed as multiple spot (dynamic competitive) markets, one market per link (similar to the New York Stock Exchange). These spot markets operate independently and asynchronously since there is no need for market communication (for example, price comparisons) or synchronization from switch to switch. Consequently, this results in a decentralized economy, where the physical failure of one switch/link does not necessarily cause failure of the entire economy.

The price computation for link i is performed at the switch, at discrete intervals of time. Denote the nth calculation instant as t_n^i and the interval of time between the calculation points t_n^i and t_{n+1}^i as the nth price interval, P_n^i. The price during P_n^i is constant and is denoted as p_n^i. The demand for bandwidth at link i is measured as the total (aggregate) traffic received at its associated output port. During the nth price interval, P_n^i, the total demand is expected to change; even so, the calculation of p_{n+1}^i will only use the demand measured at the end of the interval. For this reason, let the demand for bandwidth at link i, at the end of the nth price interval, be denoted as d_n^i. The supply of bandwidth at link i is constant and denoted as s^i.

At the end of the price interval, P_n^i, the switch updates the price of link i using a *modified* tâtonnement process. A tâtonnement process sets the price of a resource with respect to the excess demand [103]. A limitation of the tâtonnement process, in its original form, is the inability to dynamically adapt to changing network demands. Yet these dynamics, such as changing number of users and non-stationary multimedia traffic, are prevalent in current integrated service networks. To handle such dynamics, a modified tâtonnement process [36] is used to price network resources,

$$p^i_{n+1} = p^i_n \cdot \frac{d^i_n}{\alpha^i \cdot s^i} \tag{6.1}$$

where p^i_{n+1} is the new bandwidth price, p^i_n is the current price and d^i_n is the aggregate demand for bandwidth. The modified tâtonnement process adjusts the price at regular intervals, based on the demand (received traffic) and the supply. The bandwidth supply is the total bandwidth times a constant α, where $0 < \alpha \leq 1$. This modification causes the price to increase after some percentage (α) of the total bandwidth has been sold. This is evident from the equation, since the price will only increase if the numerator is greater than the denominator ($d^i_n > \alpha \cdot s^i$). When demand changes, the modified tâtonnement process can dynamically adjust the price towards the new equilibrium. An *equilibrium price* p^i_* is reached at link i when the supply equals the demand. The resulting allocation at equilibrium is Pareto-optimal and fair (chapter 5 defines fairness). This dynamic bandwidth market is also state-less since the price is calculated using only the aggregate demand, supply and current price.

After the new price, p^i_{n+1}, is calculated it is delivered to each NB using this link. A simple technique for distributing new prices is discussed in section 6.4.2, other price distribution methods are described in [38, 39]. As previously stated, the unique advantages of the dynamic competitive market include: ability to adjust to network dynamics, state-less implementation, distributed control of individual QoS, no restrictions on the statistical behavior of the users and profit maximization during "congested periods."

6.1.2 User

User j, executing a network application, requires bandwidth for transmission. The amount of bandwidth desired is determined from the application and is denoted as b^j. Assume b^j is constant for the duration of the application. In section 6.3, b^j is allowed to vary over time, which is desirable for multimedia transmission.

Based on prices and wealth, the user can afford a range of bandwidth (less than or equal to b^j), and some amounts will be preferred over others. In economics these preferences are represented with a utility function (curve), which provides an important link between resource amounts and user satisfaction. For this economy *QoS profiles* [85] are used as the utility curves. Based on psycho-visual experiments, the QoS profile is a two dimensional graph, as seen in figure 6.2. The profile can be approximated by a piece-wise linear curve with three different slopes. The slope of each linear segment represents the rate at which the performance of the application degrades when the network allocates a percentage of the desired bandwidth, b^j. A steeper slope indicates the inability of the application to easily scale bandwidth (for example, high quality video), while a flatter slope signifies the application can more readily scale bandwidth requirements (for example, teleconferencing or data transmission). The horizontal axis measures the bandwidth ratio of allocated bandwidth to desired bandwidth, b^j. The vertical axis measures the satisfaction and is referred to as a QoS score. For this these, QoS scores range from one to five, with five representing an excellent perceived quality and one representing very poor quality. Any value greater than or equal to 3 will be referred to as an *acceptable* QoS score. As seen in the figure, if the allocated bandwidth is equal to the desired bandwidth, the ratio is one and the corresponding QoS score is 5 (excellent quality). As this ratio becomes smaller the QoS score reduces as well. Profiles can be created for a variety of applications and redefined as users gain more experience. New and updated profiles can be easily incorporated within the economy as they become available. More information about QoS profiles and the relationship between bit-rate and quality can be found in [66, 74, 85].

Finally, user j is charged continuously for the duration of the session (analogous to a meter). Assume user j will provide an equal amount of money over regular periods of time to pay for expenses. The budget rate of user j will be denoted as W^j ($/sec). A single initial endowment could have been used, but would necessitate defining how it is spent during the session. To simplify simulation and analysis, budget rates are used.

6.1.3 Network Broker

A user can only enter the network economy through a network broker (NB). The NB serves as an agent for the user and is located between the user and the edge of the network. The functions of the NB can be part of the protocol stack that executes on a host system, just as current protocol stacks provide flow control to individual users. Representing

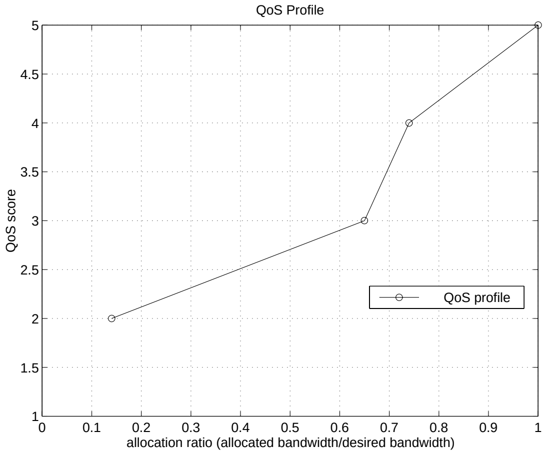


Figure 6.2: Example QoS profile.

the user in the economy the NB performs the following tasks: connection admission control, policing, and purchase decisions. Although the NB works as an agent for the user (making purchasing decisions), assume the NB operates honestly in regards to the switches and the user.

The NB controls network admission by initially requiring the user to have enough wealth to afford at least an *acceptable* QoS; otherwise, the user is denied access. The purpose of this requirement is to be certain all users are viable consumers in the market, which also prevents overloading the economy. In this thesis, it is assumed the social welfare of the economy is better when it consists of fewer users each receiving a good QoS, instead of many users each receiving a poor QoS. Hence, the goal is to maximize the number of users in the economy, where each user can afford an acceptable QoS. If the desired bandwidth is constant, then the test is relatively simple. However, for sources where the desired bandwidth will change over time, a more complex admission test is required.

The NB monitors the user and the prices by gathering and storing information about each. From the user, the NB collects and stores; the QoS profile, b^j and W^j. The NB also stores the route, R^j, that connects source to the destination. For each link in R^j, the current price p^i is collected[1]. Prices will change over time, since they represent link supply

---

[1] The requirement that the NB must know the entire route, and store a distinct price per link, can be relaxed. Methods for price distribution and collection are discussed in section 6.4.2.

and demand. For this reason, the NB will only store the most recent price from each link in the route. The NB will divide W^j into separate budget rates, one for each link in the route. Denote $w^{j,i}$ as the budget rate of user j for link $i \in R^j$. As discussed in section 5.2.1, assume for user j all $w^{j,i}$ are equal ($w^{j,h} = w^{j,i} \quad \forall h, i \in R^j$). For this reason, the second superscript of $w^{j,i}$ (i, indicating the link) will be dropped for brevity. Separate budgets are used to localize the effect of prices to each link. This prevents spending the entire budget on one expensive link. Of course depositing and withdrawing to and from these individual budgets is possible and perhaps advantageous. Using this information the NB levies the user for their consumption. Users will be charged based on usage (similar to electricity), since bandwidth is a non-storable resource. Using this information the NB polices the user, ensuring only the bandwidth purchased is used.

Finally, the NB determines the amount of dynamic bandwidth to use. This value is based on the budget, current prices and QoS profile of the user. Denote the amount of dynamic bandwidth to purchase (use) as, y^j. Once the NB determines y^j, the user will start sending at this rate immediately. There is no need for direct confirmation/feedback from the switches. A new amount of bandwidth to purchase will be determined in response to a new price (or change in demand, as will be described in section 6.3) using the following equation,

$$y^j = \min\left\{ \min_{\forall i \in R^j} \left\{ \frac{w^j}{p^i} \right\}, b^j \right\} \tag{6.2}$$

As defined by the equation, the NB uses no more bandwidth than the minimum which is affordable on any link in R^j. It is possible that y^j will be less than what is acceptable (determined from the QoS profile), due to the QoS constraint, prices and budgets. If this case arises, the user must either; increase the budget rate, accept a lower QoS, or drop the connection. This is a motivation for the second bandwidth market described in chapter 7.

As described earlier, once the NB has determined y^j it will start sending immediately at this rate. No signaling is performed. This technique provides a significant reduction in overhead; however an over allocation of resources may occur. Consider the following scenario. Assume many users are using one link and the price has reached an equilibrium value. Now assume one user exits the network and this reduction of bandwidth results in a lower price. If the remaining users react to this lower price, over-allocation of bandwidth may occur. An over-allocation may still occur if many users using a link start sending at a higher rate simultaneously due to their application (not price); however this would require a correlation of these events. In general, adjusting the price based on $\alpha \cdot s^i$ and the high capacity of most links diminish the significance of this problem.

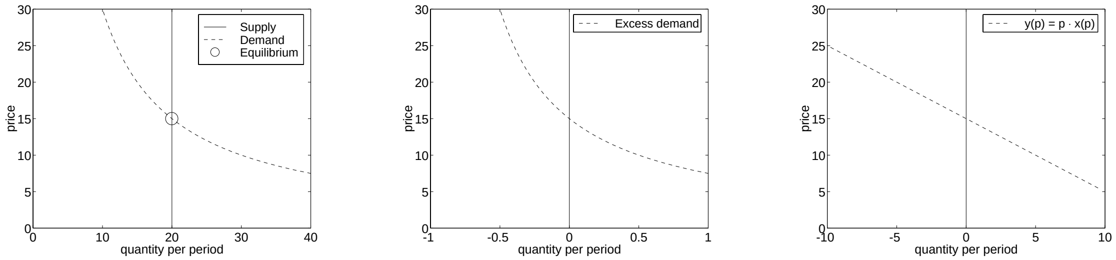
6.2 Price Stability

In section 5.2 it was proven that an economy consisting of multiple competitive markets can yield a Pareto-optimal, weighted max-min fair and equitable allocations; however, this occurs only when the market is in equilibrium (supply equals demand). For this reason, the tâtonnement process, equation 6.1, must be proven to reach the equilibrium price p_*. In this section assume the aggregate demand, d_n^i, is constant (as done in [33]). This assumption is removed in section 6.3, where the effects of network dynamics (users entering/exiting and variable user demand) are properly addressed.

The equilibrium price (p_*) occurs when a price is reached such that the demand equals the supply. At this point, the resources are fully utilized. If the demand changes, the pricing mechanism should alter the price to return to equilibrium. For that reason, adjustments in the price are driven by knowledge from the market concerning the *excess demand* at a specific price. Denote the demand for bandwidth at price p as $d(p)$. For a link in the network, the change in price over time is,

$$\frac{dp}{dt} = p \cdot \left(\frac{d(p)}{\alpha \cdot s} - 1 \right) = p \cdot x(p) \tag{6.3}$$

where $x(p)$ is the excess demand at price p. Example supply, demand and excess demand curves for the system are given in figure 6.3. As seen in figure 6.3(a), the demand curve has a negative slope which represents that an increase in price will reduce demand. The supply curve is a vertical line, because the supply of bandwidth is constant (the link does not produce bandwidth). From the supply and demand curves the *excess demand* curve can be derived. Using these graphs the behavior of the price rule 6.1 can be predicted.



(a) Supply and demand curves.

(b) Excess demand curve.

(c) Price times the excess demand ($y(p) = p \cdot x(p)$).

Figure 6.3: Example supply, demand and excess demand curves.

Define stability as,

$$\lim_{t \to \infty} p_t \rightarrow p_* \tag{6.4}$$

The price rule will increase the price p when it is lower than equilibrium price p_*. This is done because the excess demand is greater than one. When p is greater than p_*, it is lowered towards p_* because the excess demand is less than one. Therefore the price rule always moves the price towards p_*, resulting in price stability. It should be noted that the slope of the supply curve must be positive for this to be true.

The stability of the price rule can be proven mathematically as well [15]. Using the excess demand equation, the price adjustment of equation 6.1 over time can be written as,

$$\frac{dp}{dt} = p \cdot x(p) = y(p) \tag{6.5}$$

Viewing the price adjustment as a differential equation, the local response can be analyzed in the region of an equilibrium price using the Taylor approximation,

$$\frac{dp}{dt} = y(p_*) + y'(p_*) \cdot (p - p_*)$$

$$\frac{dp}{dt} = y'(p_*) \cdot (p - p_*) \tag{6.6}$$

The general solution to this equation is,

$$p = (p_0 - p_*)\, e^{y'(p_*) \cdot t} + p_* \tag{6.7}$$

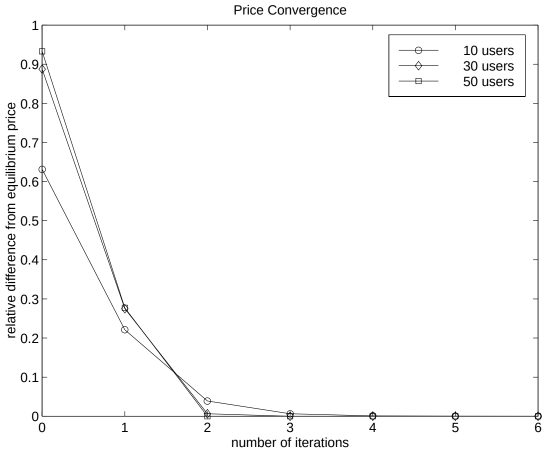


Figure 6.4: Number of iterations required for price convergence with 10, 30 and 50 users.

where p_0 is the initial price. As seen from the solution, p approaches p_* as time increases. However it must be the case that $y'(p_*)$ is negative, which is illustrated in figure 6.3(c). To provide some insight into the speed of convergence (number of iterations required by the equation), 30000 independent simulations were performed. Each consisted of a competitive market (output link) with 10, 30 or 50 users. For each simulation user j was assigned a random demand b^j and a random wealth w^j. The initial bandwidth price was 1 for each simulation. At every price iteration, the relative difference from the equilibrium price was recorded. As seen in figure 6.4, the relative difference was 0.26 after one iteration, 0.02 after two iterations and 0.002 after three iterations.

6.3 Network Dynamics and Optimality

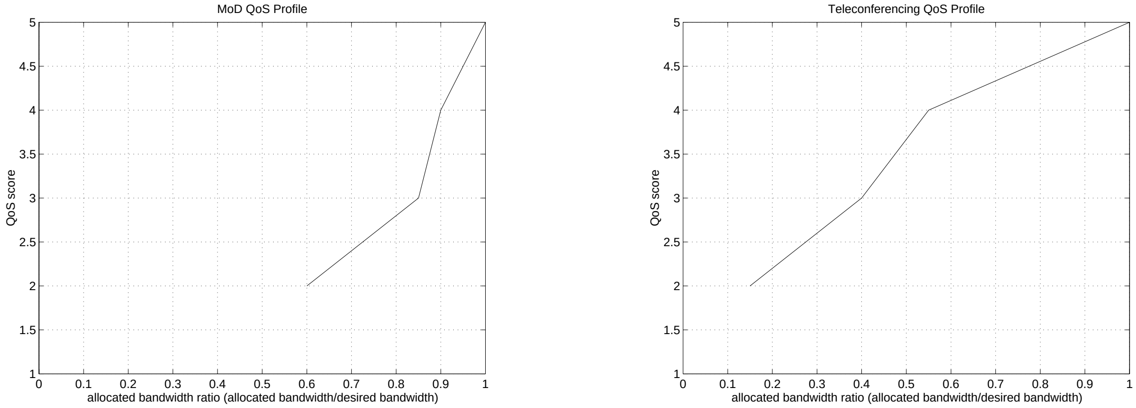
Thus far, the analysis of the network economy has not considered the dynamic nature of an actual computer network. The dynamics we are interested in include; users entering/exiting the network, and allowing Variable Bit Rate (VBR) sources. Although prevalent in actual networks, these dynamics have been either or both excluded in other microeconomic flow control methods. If the number of users and/or the demands for bandwidth change over time, then the aggregate demand, d_n, for a link will vary as well. As a result, there is not a single equilibrium price, p_*, for all time. However, the market can be viewed as having multiple equilibrium prices, each for some segment of time. During a segment the pricing technique will seek the equilibrium price as described in section 6.2. Once this price is found, the resulting distribution is Pareto-optimal and fair. When the aggregate demand changes, the stability of the price equation ensures that the price always moves towards p_*.

Due to the complexity of the changing source demands, simulations are used to demonstrate the performance of the spot market approach under network dynamics. Experimental results have shown the spot market approach achieves optimal allocations over 90% of the time under realistic network conditions [36]. Additional experimental data is presented in the next section, including a comparison with other bandwidth allocation methods.

6.4 Spot Market Experimental Results

In this section the performance of the spot market approach is investigated under network dynamics (changing user demands and users entering/exiting the network). Two sets of experiments were performed. The first set of experiments defines and tests an approximation to the wealth distribution method given in algorithm 5.1. The second set of experiments compares the performance of the spot market approach with other bandwidth allocation methods.

As described in the introduction, multimedia applications are expected to play a more prevalent role in computer networks, but are difficult to manage due to their unpredictable nature. To provide a realistic environment, each experiment simulated two different types of multimedia applications: Multimedia on Demand (MoD) and teleconferencing. MoD applications require the transmission of high quality voice and video. These applications can scale bandwidth requirements only within a limited range, since bandwidth control is achieved through quantizer control [85]. The QoS profile associated with MoD applications is given in figure 6.5(a). As seen in the profile, the acceptable bandwidth ratio range (resulting in a QoS score greater than or equal to 3) is relatively small, 0.85 to 1.0.



(a) MoD QoS profile.

(b) Teleconferencing QoS profile.

Figure 6.5: MoD and teleconferencing profiles.

Teleconferencing applications, in contrast, transmit a lower quality voice and video and can scale bandwidth requirements within a larger range. This is primarily due to quantizer control as well as the ability to transmit below the standard 24 or 30 frames-per-second. The QoS profile associated with teleconferencing applications is given in figure 6.5(b); the acceptable bandwidth ratio range is 0.4 to 1.0. Regardless of the application type, each user transmitted one of 15 MPEG-compressed video traces (described in section 3.3.2). The goal for each experiment is to achieve and equitable allocation for these two different applications.

6.4.1 Wealth Distribution Approximation

The wealth distribution algorithm for an equitable allocation (algorithm 5.1) requires complete knowledge of the users and network configuration. The availability and reliability of such information is limited in an actual network environment; therefore, a simpler approximation is needed. In this section, an approximation of the wealth distribution algorithm is presented. The performance of the approximation is compared against other wealth distributions to measure its effectiveness.

The approximation of the wealth distribution algorithm assigns wealth based on

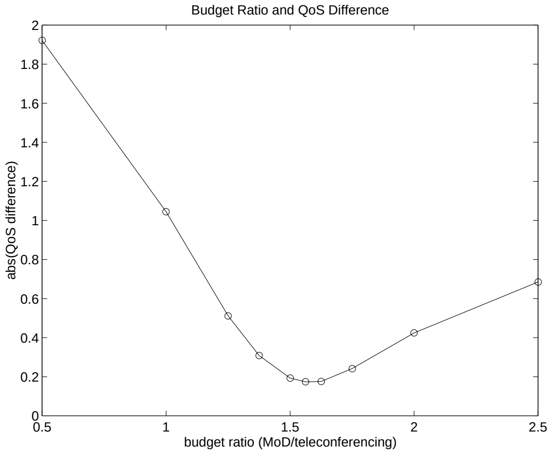


Figure 6.6: QoS difference for different budget-ratios.

the bandwidth-ratio required to achieve a QoS score of 3 for each type of application in the network [36, 38]. The ratio of these bandwidth-ratios are then used to distribute wealth for an equitable allocation. For example using the QoS profiles given in figure 6.5, a MoD user requires a bandwidth ratio of 0.85 to receive a QoS score of 3, while a teleconferencing user requires a ratio of 0.4. This yields a ratio of

$$\frac{0.85}{0.4} \approx 2$$

Assuming MoD applications have a budget rate[2] (wealth) of 3×10^8/sec, then teleconferencing users would have a budget rate of 1.5×10^8/sec.

To measure the effectiveness of this approximation, a simulation of a single 55 Mbps link with 38 users was performed. Half of the users were considered MoD applications. The remaining users were considered teleconferencing. As previously described, each user transmitted one of 15 MPEG-compressed video traces (described in section 3.3.2). As defined in section 5.2.2, an equitable allocation is one in which all users achieve the same utility. In this experiment, the difference in the QoS observed by each type of application was recorded as the budget-ratio ranged from 0.5 to 2.5 (MoD/teleconferencing). The QoS differences are given in figure 6.6. As seen in this figure, the approximation method (budget-ratio of 2) yields a QoS difference of 0.4, which is slightly higher than the lowest

---

[2] The denomination is based on bps; if based on Mbps, the budget would be 300/sec.

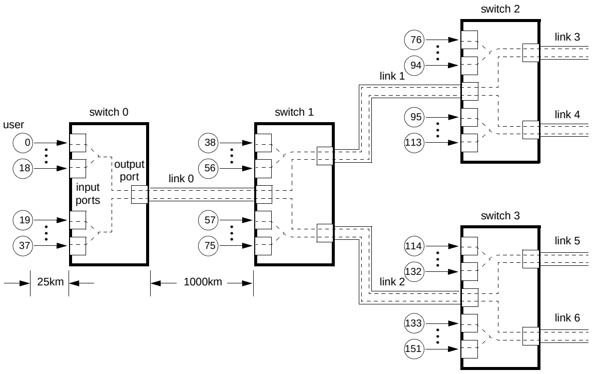


Figure 6.7: Network configuration used in the spot market comparison.

QoS difference possible (0.18). Note that a zero QoS difference is not possible, since the budget rate is constant and the source demands are variable. The approximation method is less complex than the original wealth distribution algorithm; however, it does result in an allocation slightly less equitable than possible. Considering the reduction in information required, the approximation method provides acceptable results. This is also demonstrated in the next experiment.

6.4.2 Spot Market Comparison

In this section the performance of the spot market approach is evaluated and compared with two other allocation methods: Lakshman, et al. demand-based weighted max-min [62] and a centralized max-min method. Results will show that the spot market approach achieves high network utilization and equitable (QoS-fair) allocations, as well as better QoS control than the other methods.

Comparison Configuration

Similar to [23, 62], a rate based simulator was used that propagated rate changes through the network. This resulted reduced simulation times, considering the number of users, traffic type and network modeled. The network simulated consisted of 152 users, four switches and seven 55 Mbps links, as seen in figure 6.7. The network can be described as a"parking lot" configuration, where multiple sources use a primary path. This configuration was agreed upon by members of the ATM Forum [58] as a suitable benchmark for allocation methods; it models substantial competition between users with differing routes and widely-varying propagation delays.

Half of the users (even numbered) were considered MoD, while the remaining users were teleconferencing. The application types, QoS profiles and MPEG source traces are discussed in section 6.4. Users entered the network at random times uniformly distributed between 0 and 120 seconds.

Centralized Max-Min Allocation

A comparison with max-min is provided since it is a fairness goal sought by many bandwidth allocation methods [3]. When contention occurs for link bandwidth, max-min provides each user (of the congested link) with an equal share. However, allocating equal amounts of bandwidth may not be best when considering the individual QoS expected by each user [38, 62]. The max-min implementation was centralized and no communication overhead was included for distribution allocation information. For this reason, the max-min results presented should be considered better than what is possible in practice.

Demand-Based Weighted Max-Min

In [62], Lakshman, et al. proposed an ABR explicit rate control method for transmitting compressed video. ABR explicit rate control relies on network feedback provided by Resource Management (RM) cells that are circulated for each connection [3]. The RM-cell consists of several fields, one of which is the Expected Rate (ER). This field indicates the maximum rate the network can support for this user. As the RM-cell travels along the path, a switch and/or destination may alter its contents. When a switch determines the ER of a user, it attempts to allocate the ABR bandwidth in a fair manner. The rate control method proposed by Lakshman, et al. provides a form of weighted max-min fairness. Weights are equal to the desired bandwidth of each application; therefore this method will be referred to as "demand-based weighted max-min." This method requires frame prediction to allocate bandwidth before it is required; however a look-ahead buffer was used instead. For this reason, the performance of this method should be considered best possible[3]. In addition, this method is state-maintaining since it requires per connection information to determine the allocation amount (ER). Once the cell reaches the destination it is returned to the source, which must alter transmission based on the RM-cell information.

Spot Market Approach

The spot market approach was implemented as described in section 6.1. As mentioned in section 6.1.1, there are many different methods for distributing prices. For example, the switch could send prices to users of the link, as done in [38]. Alternatively, a user could periodically generate packets (Route Price packets or RP-packets) to collect prices along the route. The purpose and function of the RP-packet is very similar to the ABR RM-cell. The RP-packet would circulate between the source and destination storing prices along the route. Switches along the route insert the current link price (on the return path). When the RP-packet returns to the source it contains an array of prices $\{p^j\}, \forall j \in R^i$; therefore, the size of the RP-packet will depend on the number of links in the route. An alternative approach would use the RP-packet to store the highest link price in the route. A switch would insert the link price only if it is higher than what is currently stored. The returned RP-cell contains "route price" for the user [39]. The single "route price" approach is used for this simulation comparison.

The following initial values were used for the spot market approach. MoD users had a budget rate, w, of 3×10^8/sec, while teleconferencing users had a budget rate of 1.5×10^8/sec. The budget rates were determined using the approximation method presented in section 6.4.1. Switches initialized their prices to 50 and α (the target utilization) to 95%. Switches updated their link prices at 10 msec intervals, a compromise between the desire for responsiveness, and the need for stability.

Comparison Results

For comparisons, the link bandwidth utilization and the QoS provided to each user were recorded. Allocation graphs are provided to measure the utilization of link bandwidth.

---

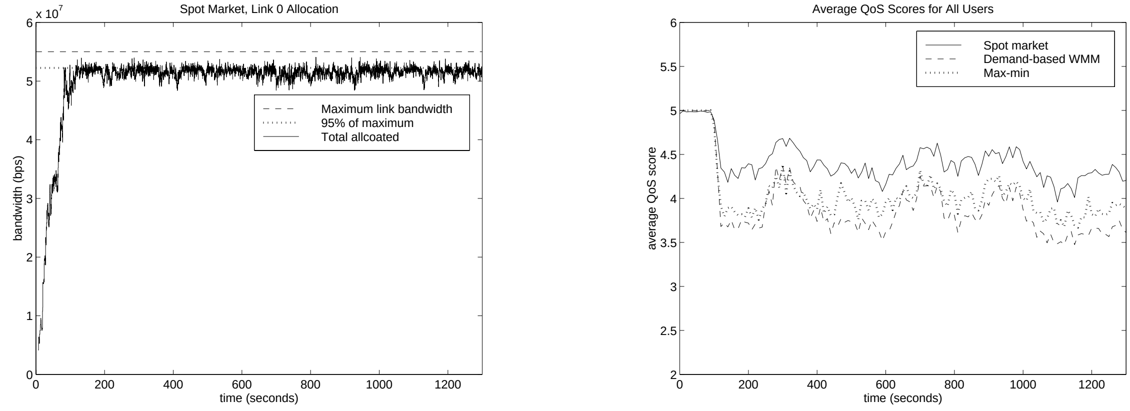
[3] A correction was made to the algorithm presented in [62] and was confirmed by the author.

To measure the QoS observed, average QoS graphs, percent Good or Better (GoB) measurements and average QoS scores are provided. Average QoS graphs measure the average QoS score observed over time and are based on all users or on individual type. The percent Good or Better (GoB) measurement is the average percentage of time a user had a quality score of at least 3.

Results from the simulation are given in figure 6.8 and in table 6.1. As seen in figure 6.8(a), the allocation provided by the spot market approach for link 0 indicates the total allocation stayed in the vicinity of 95% (α, the target utilization), yet never crossed 100%. Therefore, pricing was able to properly manage bandwidth demand (allocation results for the other links are very similar). For all users, the max-min and demand-based weighted max-min methods yielded lower average QoS and percent GoB values. This indicates, on average, users experienced lower QoS scores and enjoyed an acceptable QoS for shorter durations than the pricing method. More importantly, the pricing method provided both application types similar QoS scores and percent GoB values. This represents a more *equitable* (QoS-fair) allocation by the price method than max-min or demand-based weighted max-min. This is due to the inability of max-min or demand-based weighted max-min to differentiate between different classes of users. When equitable allocations are desired, allocation decisions must consider the fact that a reduction in bandwidth reduces the QoS for MoD users more quickly than teleconferencing users (as defined by their profiles). This was accomplished by the spot market approach via wealth distribution. The spot market approach is also state-less and requires less overhead than the other methods.

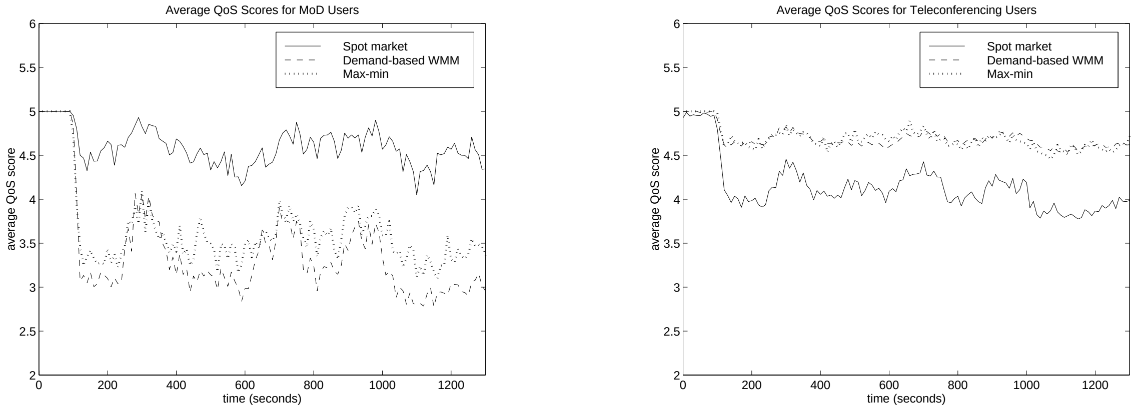
6.5 Chapter Summary

This chapter introduced a distributed, multi-user allocation method based on microeconomics. A computer network was viewed as multiple dynamic competitive (spot) markets consisting of three entities; users (those who execute network applications), Network Brokers (NB) and switches. Using competitive market nomenclature, users were consumers, switches were producers and network brokers were used to assist in the exchange of network resources. Link bandwidth was the resource exchanged in these markets, and



(a) Price ABR link 0 allocation.

(b) Average QoS score for all users.



(c) Average QoS score for MoD users.

(d) Average QoS score for teleconferencing users.

Figure 6.8: Allocation and average QoS graphs for the spot market comparison.

	%GoB			**Average QoS Score**		
	All	**MoD**	**Teleconf.**	**All**	**MoD**	**Teleconf.**
Spot market	90	90	90	4.43	4.63	4.14
Demand-based WMM	72	54	99	3.88	3.36	4.68
Max-min	80	66	99	4.25	3.92	4.76

Table 6.1: Percent GoB and average QoS scores for the spot market comparison.

is considered a non-storable resource (similar to residential electricity). Each switch owns the bandwidth of any output link connected to it and prices bandwidth based on local supply and demand. Pricing is done locally and asynchronously using a modified tâtonnement process. The modified tâtonnement process allows user demands to change over time, which is unique to the spot market approach. The user, executing a network application, requires link bandwidth for transmission and is represented in the economy with a NB. The NB, located at the edge of the network, collects prices and determines usage levels that maximize the user's utilization. Since these calculations occur at the NB, this greatly reduces the computation and storage overhead of the switches. Once a new amount is determined, it is immediately used (no signaling or reservations are required). These are unique features of the spot market approach. This spot market economy also encourages high utilization, with Pareto-optimal and equitable (QoS-fair) allocations.

An approximation of the wealth distribution algorithm (for equitable allocations, defined in section 5.2) was also defined in this chapter. The approximation greatly reduces the amount of information required to distribute wealth, that will result in an equitable allocation. Simulations demonstrate that the approximation yields comparable allocations to the original algorithm, with less source information. Simulation results also demonstrate the ability of the spot market to successfully allocate bandwidth of a network to a large number of diverse users, each transmitting an actual MPEG-compressed video trace. The economy also provided substantially better control of QoS than max-min or demand-based weighted max-min [62]. The allocation calculations are performed at the edge of the network. This reduces the computation and storage overhead of switches. The implementation cost of the spot market method will be very reasonable, since it is a state-less technique.

One disadvantage of the spot market approach is the absence of resource guarantees. It is possible that a user may enter the network during a period of time when prices are low, then at a later time find prices have increased dramatically. These unpredictable price changes can cause the QoS of the user to suffer or even force the user to exit the network prematurely. For this reason, a method is needed to provide guarantees of resource availability (price stability). The multi-market approach addresses this issue and is described in the next chapter.