Information Economics and Policy, Special Issue on Multimedia, Vol. 10, No. 1, Mar. 1998, pp. 127-55.

The Results of Competition Between Integrated-Services Telecommunications Carriers*

Jon M. Peha† and Saurabh Tewari
Carnegie Mellon University
Department of Electrical and Computer Engineering
Pittsburgh, PA 15213-3890

Abstract

This paper investigates the results of competition between two profit-seeking telecommunications carriers, as might occur when cable TV providers compete with local exchange carriers. Each firm has a fixed capacity that it can use to offer two different services, such as telephony or video. We explore whether an equilibrium exists at which no carrier has any incentive to change its prices and outputs. For arbitrary services and demand, we show that an equilibrium may not exist, or that multiple equilibria may exist which means non-market forces such as regulation and entry strategy might determine the final outcome of competition. Furthermore, it is shown that a firm may not choose to compete in the market for one of the services, and thus, the lack of market share does not imply a barrier to entry. However, we also show that these confounding phenomena are less likely to occur with the services that will probably be offered first on commercial integrated-services networks, like telephony, pay-per-view movies, and email.

Key words: competition, integrated services, oligopoly, pricing, telecommunications.

1 Introduction

It is becoming practical for telecommunications firms to offer diverse services, such as telephony, pay-per-view movies, Internet access, and videoconferencing on the same physical network. This paper investigates the results of competition between two profit-seeking telecommunications carriers, as might occur when cable TV providers compete with local exchange carriers. To simplify the analysis we restrict ourselves to the case where these two firms offer only two services. This is also of interest because firms are likely to begin competition in the well-established markets of

*This material is based in part upon work supported under Grant NCR-9210626 from the National Science Foundation.

†Corresponding author: (412) 268-7126, peha@ece.cmu.edu, http://www.ece.cmu.edu/afs/ece/usr/peha/peha.html

video and telephony. It is assumed that the firms do not collude and that capacities and demands do not vary during the period addressed in our analysis.

The paper investigates the following questions: Will the prices become stable (i.e. reach a point where no one firm can increase its revenue by changing its price for any service)? We define this point where prices become stable as an equilibrium. If prices do become stable, what factors influence the resulting prices? Are there multiple such equilibria? If multiple equilibria exist, non-market forces such as regulation and entry strategies become important in that they can influence which equilibrium is reached. Do both firms offer both services at the equilibria? If one of the firms may choose not to offer one service, then lack of market share does not imply the existence of a barrier to entry. These issues are critical to regulators and industry leaders who must manage the transition from monopoly single-service networks to competing integrated-services networks. Regulation and oversight can be eased when there is reason to believe that the firms are capable of competition, and in the mean time, neither firm should be given an artificial advantage that is sustainable when the regulation ends.

Section 2 describes our model. Section 3 shows how the nature of supply and demand determines the answers to many of the questions above. Since those answers depend greatly on supply, which is a complex commodity in integrated-services networks, we characterize the supply anticipated in realistic scenarios in Section 4. Conclusions are presented in Section 5.

2 Problem Formulation

In this section, we present the mathematical framework we use to investigate the results of competition. The general model for the telecommunications industry is described in Section 2.1. Section 2.2 refines the model for the special case where there are two firms each offering two services, and presents our basic approach to studying competition and characterizing equilibria.

2.1 The General Model

This section first defines the terms related to supply of and demand for networks services, and states assumptions. The section then defines an equilibrium, and shows some of its characteristics.

The N firms offer M different services, each of which is designed to meet the needs of different applications. For example, a service k designed for voice will provide information streams with the anticipated characteristics (e.g. data rate, burstiness) with a quality of service acceptable for telephony, as well as the acceptable blocking probability β_k. β_k is the probability that an arriving call will receive a busy signal due to lack of available resources. Selecting an appropriate β_k is part of designing the service (Wang et al., 1997). x_{ik} is the amount of service k traffic carried by firm i, as expressed in call arrival rate (including calls that will be blocked), and firm i's output vector $\vec{x_i}$ is $(x_{i1}, \ldots, x_{iM})$. The network capacity limits how large a call arrival rate of each service that a firm can handle and still meet performance requirements. The set of all such feasible output vectors for firm i is defined as the *admissible load region* for firm i and is represented by X_i, where firm i can support a load $\vec{x_i}$ if $\vec{x_i} \in X_i$. Implicit in this model is the assumption that, as long as performance requirements are met (e.g. delays and blocking probabilities are within the acceptable range for the given service), demand does not depend on performance.

The price of a service depends on the *demand function* $s_k(\vec{p})$, which describes the amount of service k demanded by the customers of all of the firms when the prices available to them are

$\vec{p} = (p_1, \ldots, p_M)^T$, where p_k is the price for service k. We assume that call duration is independent of price, as would be appropriate when customers are charged per call, or per month. Also, there is no charge for blocked calls.

We make the following assumptions on the demand functions.
Assumption 1:
(a) $s_k(\vec{p})$ does not change over time during the period of interest.
(b) There are no externalities i.e. $s_k(\vec{p})$ does not depend on current or previous output levels.
(c) The demand for service k is independent of the demand (price) for service j, $\forall\ j \neq k$, so that $s_k(\vec{p}) = s_k(p_k)$.
(d) $s_k(p_k)$ is continuous and differentiable for all k.
(e) Demand for a service decreases if the price for the service increases, i.e. $\frac{ds_k}{dp_k} < 0 : \forall\ k$.

Demands may have time-of-day fluctuations which are typically accommodated by setting different peak-hour prices and off-peak-hour prices (Wang et al., 1997). This can be done by applying the model in this paper once to establish peak-hour prices using peak-hour demand, and then again for other periods where demand is different. More generally, one might define a peak-hour call as a different service from a late-night call. (Since demands are assumed to be independent, this would mean that consumers do not shift usage between time periods.) However, prices are not allowed to change with the instantaneous fluctuations of network state, as in (Peha, 1997).

Assumption 1(c) enables two more useful definitions. First, we define the *price function* $f_i(x_i)$ to be the inverse of the demand function $s_i(p_i)$. $s_i(p_i)$ is continuous and $\frac{ds_i}{dp_i} < 0$, so $f_i(x_i)$ is a function. Assumption 1(e) implies that $\frac{df_i(x_i)}{dx_i} < 0\ \forall\ i$. Second, demand can also be expressed with a *value function* $v(p)$, which describes the arrival rate of calls with a value of p to the user. Users will pay for a service if the value exceeds the price, so $s(p) = \int_p^\infty v(a)da$.

We make the following assumption regarding consumer behavior in selecting among firms to buy a service.
Assumption 2: Customers always buy a service from the firm which charges the lowest price for the service and there are no costs for switching from one firm to another.

We make the following assumptions about the firms offering telecommunication services.
Assumption 3:
(a) Firms seek to maximize their profits.
(b) Firms do not collude.
(c) All the firms know that all other firms intend to maximize their profit, and have complete information about all demand functions and admissible load regions.
(d) Capacity costs are sunk, and any costs that depend on the amount of traffic carried are negligible in comparison.
(e) Combined installed capacity of all the firms is such that there exists a set of non-zero prices for which the total demand will exceed the capacity of all firms combined.
(f) Firms use all their capacity.

Assumption 3(f) implies that at equilibrium, the firms operate on the boundary of the admissible load region. Although it is not always valid, it is generally reasonable for two reasons. First, competitive pressure can force the firms to use all their capacity at equilibrium, as suggested by Bertrand. If one firm had free capacity, it could slightly undercut its competitor's price, thereby using all its capacity and increasing its revenue. The competitor may do the same until both carriers use all their capacity. (If the carriers have built too much capacity, such price competition could even lead to revenues that fall below costs. Competition is not sustainable in this scenario, but that's beyond the scope of this paper.) Second, capacity is expensive and firms in a stable unregulated industry are unlikely to build capacity that they don't expect to use, at least in the peak hours. In fact, for peak-hour pricing, the marginal revenue derived from network capacity should be positive when all capacity is being used to justify building that much capacity. (Although there may be a transient period soon after the infrastructure is created/expanded to support a new service when this is not the case.) Indeed, some researchers have viewed this process as a two-stage game in which both prices and capacity levels are set. This paper focuses only on the latter part of the game, in which capacities are fixed.

Ultimately, we are interested in equilibrium prices and outputs, where an equilibrium is a point where no firm can increase its revenue by changing its price for any service. Equilibria in this model have the following two characteristics. Proofs for these, and all other propositions in this paper, are in the appendix.

Proposition 1: Under assumptions 1, 2, and 3a-d, at equilibrium, all the firms offer each service at the same price.

Proposition 2: Under assumptions 1, 2, and 3, at equilibrium, the price for a service is equal to the value of the least valuable traffic carried by any firm.

These propositions hold for any number of services, and when there is only one service, these propositions reduce to established results (Daugherty 1988, Davidson and Deneckere 1986, Friedman 1983, Herk 1993, Kreps and Scheinkman 1983).

2.2 The Two-Firms Two-Services Case

In this section, we refine the model presented in Section 2.1 for the special case where there are two firms each offering two services, and present our approach for the study of equilibria. By definition, at equilibrium, neither firm has an incentive to change its prices or outputs. Therefore, a necessary condition for equilibrium (unless one of the outputs is 0) is: for i=1, 2

$$\frac{dR_i}{dp_{i1}} = \frac{dR_i}{dx_{i1}}\frac{dx_{i1}}{dp_{i1}} = 0$$

By assumption 3(f), each firm operates on the boundary of its admissible load region at equilibrium. With only two services, this boundary can be expressed as follows. Let $a_i(x_{i1})$ be the maximum amount of service 2 that firm i can carry while the amount of service 1 it carries is x_{i1}. At the boundary of this region, at least one service j must be experiencing the maximum tolerable blocking probability β_j, and typically both will, which is what we will assume here. (If service 1 blocking probability is at its maximum and service 2 blocking probability is not, then the network could carry more of both if it simply reserved a bit more bandwidth for service 1 traffic.) Since demand for the two services is independent and both firms will choose the same prices, the revenue R_i of firm i at equilibrium is

$$R_i = f_1(x_{11} + x_{21})B_1 x_{i1} + f_2(a_1(x_{11}) + a_2(x_{21}))B_2 a_i(x_{i1})$$

where $B_j = 1 - \beta_j$ is the fraction of class j calls that are not blocked when operating at the edge of the admissible load region.

$\frac{dx_{i1}}{dp_{i1}}$ must also be characterized to derive equilibrium conditions. We know from Proposition 1 that all firms will charge the same price for each service, so when one firm changes its price, the other follows. We assume that the firm that changed its price first is the one whose output will change (Kreps and Scheinkman 1983). As one simple example, if firm 1 lowers its price, all the new customers that subscribe due to the lower price go to firm 1, and firm 2 matches the price in time to prevent its current customers from defecting. Thus, $\frac{dx_{ij}}{dp_{ij}} = \frac{ds_i}{dp_i}$, which is non-zero by assumption 1(e). Thus, $\frac{dR_i}{dx_{i1}} = 0$ at equilibrium. The necessary conditions can be written as: for i=1, 2

$$\frac{dR_i}{dx_{i1}} = B_1 x_{i1} \frac{df_1(s_1)}{ds_1} + B_1 f_1(s_1) + B_2 a_i(x_{i1}) \frac{df_2(s_2)}{ds_2} \frac{da_i(x_{i1})}{dx_{i1}} + B_2 f_2(s_2) \frac{da_i(x_{i1})}{dx_{i1}} = 0$$

where $s_1 = x_{11} + x_{21}$, and $s_2 = x_{12} + x_{22} = a_1(x_{11}) + a_2(x_{21})$. If at equilibrium, $x_{i1} = 0$, the equivalent condition would be $\frac{dR_i}{dx_{i1}} \leq 0$.

$\frac{dR_i}{dx_{i1}} = 0$ defines the *reaction function* for firm i which describes firm i's revenue-maximizing output as a function of the rival firm's output, where revenue maximization is subject to the constraint that the prices are equal. The intersection of reaction functions of the two firms is clearly a necessary condition for equilibrium. This condition is also sufficient, as shown in Proposition 4.[1] Consequently, the focus of our work is on the characteristics of reaction functions.

Proposition 3: Under assumptions 1, 2, and 3, if the system can produce a given set of prices and outputs $x_{11}, x_{12}, x_{21}, x_{22}, p_1, p_2$, it can do it in such a way that the value of the least valuable class i traffic carried by firm 1 equals the value of the least valuable class i traffic carried by firm 2.

Proposition 4: Under assumptions 1, 2, and 3, an intersection point for the two reaction functions must be an equilibrium point.

3 Characterization of Equilibria

In Section 3.1, we addresses the characteristics of reaction functions. Section 3.2 discusses the existence and stability of equilibrium points. Multiplicity of equilibria is investigated in Section 3.3, and it is also demonstrated that two carriers with identical infrastructure may end up with different revenue. Section 3.4 discusses the circumstances where one firm may choose not to compete in the market for one service. The results of Section 3 are summarized in Section 3.5.

3.1 Reaction Functions

We will concentrate on two characteristics of reaction functions: continuity and monotonicity. As will be shown in the following sections, these two characteristics are major factors influencing the results of competition. Intuitively, one might expect that a small change in one firm's strategy would lead to a small adjustment in the competitor's strategy, and that if one firm chooses to carry more of a service, its competitor would choose to carry less. This would lead to a reaction function that is continuous and monotonically decreasing. However, intuition can be wrong. We discuss continuity first.

Firm 1's reaction function gives the revenue-maximizing output for firm 1, given the output of firm 2. As shown in Section 2.2, all points on the reaction function (with the possible exception of $x_{11} = 0$) satisfy the condition $\frac{dR_1}{dx_{11}} = 0$. However, all points satisfying this condition need not be

[1] Note that Propositions 3 and 4 depend on assumptions 1(d) and 1(e). If neither of these assumptions were valid, then it is theoretically possible that there could be scenarios in which reaction functions intersect but prices oscillate without ever reaching an equilibrium.

on the reaction function. The firms maximize the revenue *globally* whereas $\frac{dR_1}{dx_{11}} = 0$ is a condition for a *local* maximum. For a given value of x_{21}, firm 1's revenue R_1 may have several local maxima over x_{11}, and this can cause discontinuities. As the proposition below shows, under the reasonable assumption that admissible load regions are differentiable, if there is one local maximum, then the reaction function must be continuous.

Proposition 5: Under assumptions 1, 2 and 3, if R_1 has only one local maximum over x_{11} for every x_{21}, and the admissible load regions $a_1(x_{11})$ and $a_2(x_{21})$ are differentiable, then firm 1's reaction function is continuous in x_{11}.

However, if firm 1's revenue can have multiple local maxima over x_{11}, then there could be discontinuities in a reaction function. This is illustrated in Figure 1. In this example, there are two local maxima for x_{21} around $x_{21} = 0.64$. Consequently, as x_{21} increases beyond 0.64, Firm 1's best response to x_{21} drops abruptly from a positive value near 0.24 to zero, thereby creating a discontinuity in the reaction function. As will be shown later in this section, this significantly affects the kinds of equilibria that can occur. Since discontinuities are impossible when there is only one local maximum, Proposition 6 presents conditions for continuity (where x_{i1}^{max} is the maximum amount of service 1 that carrier i can support when it carries no service 2 traffic).

Proposition 6: Under assumptions 1, 2 and 3, and if, for $0 < x_{11} < x_{11}^{max}$ and $0 < x_{21} < x_{21}^{max}$:
(i) $\frac{d^2 f_1(x_1)}{dx_1^2}, \frac{d^2 f_2(x_2)}{dx_2^2} \leq 0$,
(ii) $\frac{d^2 a_1(x_{11})}{dx_{11}^2} \leq 0$, and
(iii) $\frac{dR_{11}}{dx_{11}} \geq 0$ or $\frac{dR_{12}}{dx_{12}} \geq 0$ over all x_{11},
then R_1 has only one local maximum over x_{11} for every x_{21}.

For condition (iii) to be untrue, there must be so much capacity that both firms could increase revenue by leaving some capacity idle. It is unlikely that firms would build more capacity than they would want to use, at least in the peak hours when the majority of the revenue is generated. Indeed, with such excess capacity, there will be strong motivation for firms to leave some capacity idle, probably through some form of collusion.

With excess capacity, carriers must sometimes choose one of the two services and drive down that service's price so that significant revenue is still generated from the other service. This yields two possible strategies, and possibly two local maxima. Thus, it is reasonable that this condition should be stipulated to guarantee a single local maximum. The other conditions are also to be expected. When condition (ii) is violated and the admissible load region is strongly convex, it is better for a network to specialize in one of the services, again yielding two choices and possibly two local maxima. Finally, condition (i) is certainly not surprising, since there can be multiple local maxima even in a single-service network unless demand is concave. (As an extreme example, if output is inversely proportional to price, then every price yields maximum revenue.)

We now discuss the monotonicity of reaction functions.

Proposition 7: Under assumptions 1, 2 and 3, if the reaction function is continuous and the price functions are concave, then the reaction function is monotonically decreasing.

Otherwise, if the price function is convex, the reaction function could be non-monotonic, as shown in Figure 2.

Thus, in the absence of excess capacity, linear or concave admissible load regions and price functions imply a continuous and monotonically decreasing reaction function.

3.2 Existence of Equilibria

As discussed in Section 2.2, equilibria can only occur at an intersection point between the reaction functions, and such intersections can be expected to be equilibria.

Proposition 8: Under assumptions 1, 2, 3, and if, *either*
(i) the reaction functions of both firms are continuous, *or*
(ii) the reaction function for one firm is continuous and both firms have either a monotonically increasing or a monotonically decreasing reaction function,
then at least one equilibrium exists.

With linear or concave admissible load regions and price functions, reaction functions are continuous and monotonically decreasing for peak-hour prices. From proposition 8, under these conditions, we can expect at least one equilibrium. However, as shown in Section 3.1, discontinuous and non-monotonic reaction functions are also possible. With such reaction functions, there may or may not be an equilibrium, and if there isn't, prices could fluctuate wildly. For example, Figure 3 shows a pair of reaction functions which are not continuous or monotonic (the solid lines), and there is a unique equilibrium. However, by simply shifting one of these reaction functions (the dashed line), we can produce reaction functions that do not intersect, so there is no equilibrium.

3.3 Multiple Equilibria

We presented an example in Figure 1 where there were multiple equilibria. The existence of multiple equilibria is significant because it implies that demand and supply alone do not determine the outcome of competition for some scenarios. Other factors such as regulation and entry strategies may determine which equilibrium is reached. In this section, we will show that even for a simple scenario where both firms have identical admissible load regions, a number of different outcomes are possible, and only under special circumstances does a unique equilibrium exist. Where multiple equilibria can exist, we also show that building the same infrastructure as one's competitor does not imply that equal revenues will follow.

In the example of Figure 1, the two firms have identical admissible load regions, but there are equilibria at which the firms have different outputs, and consequently different total revenues. We call these equilibria where the firms have different outputs, *unequal-output* equilibria. The equilibria at which firms carry an equal amount of services are called *equal-output* equilibria.

Proposition 9: Under assumptions 1, 2 and 3, if the firms have identical admissible load regions, and the reaction functions are continuous, then an equal-output equilibrium always exists.

There will be only one equal-output equilibrium if the reaction functions are monotonic. However, if the reaction functions are non-monotonic, reaction functions may cross the $x_{11} = x_{21}$ line any odd number of times (unless $x_{11} = x_{11}^{max}, x_{21} = x_{21}^{max}$ is also an equilibrium). Thus, there may be multiple equal-output equilibria, as in the example of Figure 2. Also note that, although there must be at least one equal output (and equal revenue) equilibrium, that equilibrium may be unstable, where an unstable equilibrium is one where any perturbation in prices or outputs will cause the system to move to another equilibrium (Novshek 1985, Tirole and Fudenberg 1991).

Proposition 10: Under assumptions 1, 2 and 3, if the firms have identical admissible load regions and the reaction functions are continuous, unequal-output equilibria can occur only in symmetric pairs.

Thus, with half the unequal output equilibria, firm 1 has the larger revenue, and with the other half, firm 2 has the larger revenue. Consequently, when multiple equilibria are possible, there is likely to be much jockeying for position at the onset of competition. However, the situation may be considerably simpler if the boundary of the admissible load region is linear or concave.

Proposition 11: Under assumptions 1, 2 and 3, if the firms have identical admissible load regions, and the boundary of the admissible load region $a(x_{i1})$ is linear or concave, then unequal-output equilibria do not exist.

This implies the following corollary, if the reaction functions are monotonic.

Corollary 1: Under assumptions 1, 2 and 3, if the firms have identical admissible load regions, the boundary of the admissible load region $a(x_{i1})$ is linear or concave, and the reaction functions are continuous and monotonic, then there is a unique equilibrium.

3.4 Zero-Output Equilibria

At two of the equilibria in Figure 1, one of the services is being offered by only one firm. We call the equilibrium where one firm does not carry one service a *zero-output* equilibrium. These equilibria prove an important point: lack of market share does not imply an asymmetric barrier to entry. In the example in Figure 1, one firm has chosen not to compete in the market for one of the services. This is significant in light of the opinion in certain quarters that the local exchange carriers should be allowed in other businesses only after they lose a certain amount of market share in their core business. If zero-output equilibria are possible, other factors must be considered when determining whether competition is possible. However, in some markets, such zero-output equilibria cannot occur.

Proposition 12: Under assumptions 1, 2 and 3, if the firms have identical admissible load regions, and *either*
(i) $a(x_{i1})$, the boundary of the admissible load region, is linear or concave, *or*
(ii) $\frac{dR_{12}}{dx_{12}} \leq 0$ for all x_{11},
then zero-output equilibria do not exist.

3.5 Summary of Equilibria Results

We have seen that in the absence of excess capacity, much is known about a system with linear or concave demand and admissible load region boundaries: reaction functions are continuous and monotonically decreasing, and an equilibrium always exists. Furthermore, in the case addressed in Sections 3.3 and 3.4 where both firms have identical infrastructure, there is a unique stable equilibrium, and neither unequal output nor zero-output equilibria are possible. However, with excess capacity, convex demand, or convex admissible load region boundaries, reaction functions can be discontinuous and non-monotonic. Furthermore, even if the reaction functions are continuous, there may be multiple equilibria, unequal-output equilibria, and zero-output equilibria.

Clearly, much depends on the shape of the demand curve and the admissible load region. We do not presume to know the demand function for current and future telecommunications services, although it is common for economists to assume that demand functions are concave. However, little thought has been given to the actual shape of admissible load regions. We address this critical issue in the next section.

4 Admissible Load Regions

This section discusses admissible load regions and the factors that determine them. It will argue that in the short term, admissible load regions are likely to have concave or linear boundaries, but in the long term, there are a wider range of possibilities.

In the case of a network designed for a single purpose, e.g. to carry telephone calls from Pittsburgh to Chicago, calculating the largest output that the network can support is relatively simple. As mentioned in Section 2.1, we measure this output in call arrival rate. Given the capacity and the voice quality desired, it is possible to determine the maximum number of simultaneous phone calls that the infrastructure can carry. When those resources are insufficient, calls must be blocked, and customers will not tolerate a blocking probability that exceeds some limit. Hence, the maximum output of this system is the call arrival rate at which the maximum tolerable blocking probability is experienced in a system which can support the given number of simultaneous calls. Any output below this maximum is within the network's admissible load region.

The case of an integrated-services network is more complex. Here, the maximum number of class j calls that can be carried depends on the capacity available for class j calls, which is a function of the number of calls from other traffic types, and the technical approach used to allocate resources among the different traffic types. Determining the maximum output here is analogous to the single-service network case, but instead of describing the number of simultaneous calls and the call arrival rate as scalar values, they must now be described as vectors. Let $\vec{n} = (n_1, \ldots, n_M)$ describe the current state of the system, where n_i is the number of class i calls in progress, and $\vec{x} = (x_1, \ldots, x_M)$ describe the total output, where x_i is the amount of class i traffic supported as expressed here in arrival rate. If $\vec{n}$ calls can be carried simultaneously in a given network, then $\vec{n}$ is within that network's *schedulable region* (Hyman et al 1991, Peha 1993, Peha and Tobagi 1996a). Given this limit on the number of calls in progress, if the arrival rates described by $\vec{x}$ can be supported while every service experiences a blocking probability that is considered tolerable for that service, then $\vec{x}$ is within the *admissible load region* (Hyman et al 1991). The boundaries of these two regions depend on the network capacity, the nature of the services, and the technical approach used to share resources among different traffic types.

In general, a concave boundary for the schedulable or admissible load region implies a system that is more efficient when it carries multiple traffic types (i.e. there is an economy of scope), while a convex boundary implies the opposite. We will first discuss the range of schedulable regions that can be anticipated in integrated services networks, and then describe the resulting admissible load regions.

The simplest approach to resource sharing is *circuit switching*, which has long been used in telephone systems. With this approach, before communications can begin, a circuit is set up between the two telephones, and dedicated resources are reserved along the circuit's path. These resources are sufficient to meet the peak demands of the communicating parties, and no one else can use these resources until the call is terminated and the circuit is released. In an integrated-services circuit-switched network, there may be several possible data rates for a circuit. For example, one video circuit may consume the same resources as twenty voice circuits. The resulting schedulable region has a roughly linear boundary. To be more precise, there can be only an integral number of class i circuits in the system, so the current load must be a multiple of the data rate of a class i stream. Thus, the schedulable region may look more like a staircase with the number of steps along the class i axis equal to the maximum number of class i calls that can be carried at one time. If no call requires a large portion of the capacity, the boundary of the schedulable region will still appear linear. However, as shown in Figure 4, for extremely-high-data-rate applications, the staircase effect becomes significant. The impact of these applications on other traffic is particularly great when the ratio of class 1 data rate to class 2 data rate is large.

The principle alternative to circuit switching is *packet switching*. Information is broken up into pieces and combined with control information to form self-contained packets. These packets are independently routed through the network. This has many advantages in an integrated-services network. First of all, it is well suited to applications that do not produce information at a constant rate, because it allows sharing. When one caller is not transmitting packets, the resources that would sit idle in a circuit-switched network can instead carry packets from another stream. This is important for emerging variable-bit-rate video systems, and computer communications which tend to transmit sporadically in large bursts. Also, while circuit-switched systems typically offer a discrete number of data rates, packet-switched networks can efficiently handle streams of any data rate. Consequently, some emerging integrated-services networks are expected to adopt a packet-switched approach known as *asynchronous transfer mode* (ATM).

The use of packet-switching introduces new complexities, which have significant impact on the schedulable region. Since resources are shared with the expectation that all streams will rarely flow at maximum rates simultaneously, there will occasionally be times when not all of the arriving traffic can be carried. Packets can then be delayed, or even dropped due to buffer overflow. Each application can tolerate different levels of delay and packet loss, and many require a priori guarantees by the network that certain performance requirements will be met. For example, video packets typically have deadlines, and packets delayed beyond that deadline are received too late for play back. Packets exchanged in a distributed computation typically have a performance requirement based on mean packet delay. To meet such diverse requirements, a variety of algorithms and protocols are needed to regulate the flow of traffic, including algorithms to block new calls from the system to keep load sufficiently low, to order the transmissions of queued packets, to select packets to be dropped upon buffer overflow, and to reduce the flow of traffic into the network when it becomes congested. Any of these algorithms may or may not discriminate among packets and calls based on their performance requirements. There is no consensus in the networking community yet as to which set of algorithms should be deployed.

In a packet-switched network, a schedulable region describes capacity at the packet level. For example, it is affected by the average number of packets per second arriving in a class i call, once that has begun, as well as the burstiness of packet arrivals. The schedulable region is also affected by the users' requirements regarding packet delay and packet loss, but the likelyhood that calls will arrive or depart is irrelevant. In contrast, the admissible load region describes capacity at the call level. In addition to the factors above, the admissible load region is affected by the arrival rate of calls, and the burstiness of those call arrivals. It is affected by call duration. It is also affected by performance at the call level, which typically means call blocking probability. These two methods of representing network capacity are powerful tools for economic analysis.

Once the services have been defined, the traffic control approach selected, along with the system capacity (bandwidth, buffer space, processing capability) determine the schedulable region. More sophisticated traffic control approaches are likely to yield a larger schedulable region (Peha 1995), but the approach can also affect the region's shape. Some approaches such as the Priority Token Bank (Peha 1993, Lynn and Peha 1997), Cost-Based Scheduling (Peha and Tobagi 1996a), and Magnet Real-Time Scheduling (Hyman et al. 1991), are designed to tailor the performance of different streams to their respective performance objectives. Consider video traffic for which maximum delay should not exceed 30 ms, and computer data bursts for which mean queueing delay should not exceed 5 ms. These algorithms would meet the needs of video packets by giving them higher priority when they are in danger of missing their deadlines, which is relatively infrequent, and would otherwise give higher priority to the data bursts to improve their mean delay. When such approaches are used, the network is more efficient when carrying a mix of traffic with very different performance requirements. Figure 5 shows the schedulable regions with variable bit-rate (VBR) 1 Mb/s video and data bursts. Results were achieved through simulation using techniques described in (Peha 1996b), and the model from Maglaris et al (1988). The 95% confidence interval for these simulation results is within 5% of the value shown. The schedulable region of sophisticated approaches like the Priority Token Bank and Cost-Based Scheduling have roughly concave boundaries. In contrast, polling-based schemes such as (Demers et al 1989, Golestani 1991, Kalmanek et al 1990, Parekh and Gallager 1994), which are variations of round-robin scheduling, yield a roughly linear schedulable region.

While mixing traffic with dissimilar performance requirements can lead to a concave boundary for schedulable regions, at least when sophisticated traffic control approaches are used, mixing traffic that is generated at a relatively constant-bit-rate (CBR) with highly bursty traffic can have the opposite effect. Figure 6 shows the schedulable regions with round-robin scheduling for CBR traffic and variable-bit-rate (VBR) traffic with different burstiness. Schedulable regions are shown for different average data rates. For sources with negligible data rates, statistical multiplexing eliminates the burstiness, and the boundary of the schedulable region is linear. For traffic with higher burstiness, this boundary is slightly convex. This can be explained as follows. Throughput declines when the aggregate traffic is more bursty. Burstiness can be eliminated by carrying only CBR traffic. If some VBR traffic is supported, then the more one carries, the more the aggregate traffic is smoothed due to statistical averaging. This generally produces a slightly convex boundary. As already discussed, at high source rates, the boundary of schedulable regions looks like a staircase superimposed on the linear or slightly convex boundary for the schedulable region. Similar results have been reported by Lee and Mark (1995) for a finite-buffer system with two traffic types, and by Sidhu and Jordan (1995).

Thus, the boundary of schedulable regions can be convex, concave, linear, or staircase versions of any of these, depending on the service in question, and on the technical approach. We now consider the implications for admissible load regions. Admissible load regions are determined from the schedulable region and the tolerable blocking probabilities. We will first consider the impact of the schedulable region when all services require the same blocking probability. Unequal blocking probabilities will then be addressed.

We begin with linear schedulable regions. Schedulable regions with linear boundaries generally yield admissible load regions with linear boundaries. If the data rates of various streams are comparable, or small compared to the link data rates, then the admissible load regions are exactly linear. Only if the staircase effect is severe, i.e. if one class of traffic has an extremely high data rate, and the other is much smaller, does this significantly affect the admissible load region. In this case, the admissible load region also takes on some of the properties of a staircase, the boundary of which has both concave and convex regions. This can be seen in the examples in Figure 7. In both the examples, circuit switching is used so that the boundary of the schedulable regions is roughly linear. The figure shows the admissible load region for two scenarios. The axes are expressed in terms of average data rate from a service, where average data rate is the product of call arrival rate, average call duration, and average data rate per call. The network consists of a single 150 Mb/s link, as could be appropriate for ATM networks. In one scenario, 50 Kb/s calls, representative of constant-bit-rate voice, and 5 Mb/s calls, are carried. Even this 100 to 1 ratio yields a roughly linear admissible load region except at the edges. In contrast, in the other scenario, 50 Kb/s calls and 30 Mb/s calls, the latter of which is representative of peak-rate high definition television (HDTV), are carried. This yields a staircase-like boundary of the admissible load region. Thus, the data rates must be extremely high for this staircase phenomenon to occur.

For non-linear schedulable regions, unless the difference in data rates of the various traffic types is extreme and one stream can consume much of the network resources, an admissible load region closely mirrors the corresponding schedulable region. Indeed, in cases where the maximum number of class i calls that a link can carry simultaneously is on the order of the inverse of the desired class i blocking probability, the shapes of the schedulable and admissible load regions are visually indistinguishable. This is not an unusual condition. For example, it is true of the example in Figure 5.

The final way that different services can differ is in the blocking probabilities. For example, telephone blocking probabilities should not exceed 1%. For some applications, like 911 calls, a lower blocking probability is desirable. In contrast, pay-per-view movie customers would probably tolerate a few minutes of blocking before their call goes through. If the difference in blocking probabilities is great enough, it can make the admissible load region boundary concave. Figure 8 shows the admissible load region when two services are identical except for the blocking probability requirements. One service has very strict blocking probability requirements. The system achieves this strict blocking probability requirement by reserving some capacity for the more important calls only. We vary data rates from negligible to 15 Mb/s in a 150 Mb/s channel. When the data rate is so great that only a small number of simultaneous calls can be supported, however, there are noticeable transitions as the capacity held in reserve goes from one call's worth to two.

So what does this all mean in practice? In the short term, admissible load regions are likely to have roughly linear or concave boundaries. Voice and video are likely to be early services. Since they have similar performance requirements, and are not extremely bursty or extremely high data rate, and may have different blocking probabilities, a linear or concave boundary is likely. Although the results of competition with more than two services are beyond the scope of this paper, computer communications (e.g. Internet access) also constitute an important set of services, and their performance requirements will differ greatly from voice or video. Services for individual consumers are initially likely to be relatively low data-rate. Business applications, e.g. connecting local-area-networks in an enterprise network, will have greater data rates, but early applications will probably have lax performance requirements, making it possible to smooth traffic by delaying transmissions. Thus, we expect that the dissimilar performance requirements will lead to concave or linear boundaries. Any differences in the blocking probability requirements can only make the boundary more concave.

In the long term, this may not remain true. To yield a convex boundary to the admissible load region, or a staircase, traffic with significant performance requirements must be high data-rate, and probably highly bursty. Such applications may very well become important, such as high definition television (HDTV), browsing of high-resolution images, or multimedia collaborative work tools.

5 Conclusions

This paper investigates the results of competition between two profit-seeking telecommunications carriers who offer two services each, as may be relevant when cable TV providers compete with local exchange carriers. It was assumed that the firms have a fixed capacity and do not collude. The demand for the two services was assumed to be independent and devoid of externalities. Consequently, this model does not yet address the myriad complexities of the emerging telecommunications industry. While further research is clearly needed that broadens these assumptions, the results of competition in this case are still instructive.

The feasible output of a firm depends on the services it carries, the technical approach it uses to allocate resources between the services carried, and the network capacity. Consequently, the outcome of competition also depends on these factors, as well as the demand. In general, an equilibrium is not guaranteed for all kind of services. It is shown that multiple equilibria are possible and hence non-market forces such as regulation and entry strategy might determine the final outcome, causing firms to jockey for initial position. Furthermore, it is shown that a firm may not choose to compete in the market for one of the services, and thus, the lack of market share does not imply a barrier to entry. The potential existence of multiple equilibria, and of equilibria where one firm chooses not to compete in one service, greatly complicate the task of managing the transition from monopoly single-service networks to competing integrated-services networks.

To understand this transition, we focus on services that are likely to be offered initially by integrated-services networks. We have argued that their admissible load regions are likely to have nearly linear or concave boundaries. In this case, if the price functions are concave, an equilibrium will always exist for peak-hour prices. Furthermore, if the firms have identical admissible load regions and the price functions are concave, there is a unique equilibrium for peak-hour prices. At this unique equilibrium, the firms make the same revenue, and both firms will choose to offer both services. Thus, for example, cable companies will choose to offer voice services as soon as they are capable. If our results prove valid in other scenarios e.g. even when the demands for the services are not independent, and possibly with more than two services and more heterogeneous admissible load regions, transition to a competitive industry structure may be greatly simplified.

A Appendix - Proofs

Proposition 1: Under assumptions 1, 2, and 3a-d, at equilibrium, all the firms offer each service at the same price.

Proof: We will first show by contradiction that if all the firms that carry a service do not offer the service at equal price, a firm or a set of customers will have incentive to change strategy. Assume the system is in equilibrium. Without loss of generality, let firm 1 be the firm with the lowest price for service 1. There are two possibilities: firm 1 has idle capacity, or firm 1 has no idle capacity. We address each case in turn. If firm 1 has any idle capacity, the customers of firms other than firm 1 will switch to the less expensive firm 1, contradicting our assumption that the system was in equilibrium. If firm 1 has no capacity idle, firm 1 can increase its price by a small amount so that its price is still lower than other firms. It may or may not lose customers. If no customer leaves, firm 1's revenue increases. If some customers quit, firm 1 will have idle capacity and customers from other firms will switch to firm 1, and firm 1's revenue will still increase. Thus revenue was not maximized, contradicting our assumption that the system was in equilibrium. For the firms that do not carry a service, we can still say that their price for the service is equal to the price at which other firms offer the service. *Q.E.D.*

Proposition 2: Under assumptions 1, 2, and 3, at equilibrium, the price for a service is equal to the value of the least valuable traffic carried by any firm.

Proof: We will show by contradiction that if the price for a service is not equal to the value of the least valuable traffic carried by any firm, a firm or a set of customers will have incentive to change strategy. Assume that the system is in equilibrium. As proven above, firms must have identical prices for all services. Assume that the price does not equal the value of the least valuable traffic carried. If this price exceeds the minimum value, some customers have paid more for the service than it is worth to them, contradicting our definition of value. If the minimum value exceeds price, then firms can increase price without losing any customers. Thus revenue was not maximized, contradicting our assumption that the system was in equilibrium. *Q.E.D.*

Proposition 3: Under assumptions 1, 2, and 3, if the system can produce a given set of prices and outputs $x_{11}, x_{12}, x_{21}, x_{22}, p_1, p_2$, it can do it in such a way that the value of the least valuable class i traffic carried by firm 1 equals the value of the least valuable class i traffic carried by firm 2.

Proof: For these prices and outputs to be feasible, $p_i = f_i(x_{1i} + x_{2i})$. By assumption $1(d)$, the demand function is continuous at price p_i. By assumption $1(e)$, the amount of traffic $v_i(p_i)dp$ with a value within $[p_i, p_i + dp]$ is > 0, for a negligible dp. If both firms carry class i traffic (i.e. $x_{1i} > 0, x_{2i} > 0$), then it is always possible for this amount of traffic $v_i(p_i)dp$ to be distributed across both firms. If a firm carries no class i traffic, then without loss of generality, we can define the value of the least valuable class i traffic that the firm carries to be the same as that of the other firm. *Q.E.D.*

Proposition 4: Under assumptions 1, 2, and 3, an intersection point for the two reaction functions must be an equilibrium point.

Proof: An intersection point identifies a set of outputs where neither firm can improve its own revenue by changing its own price, assuming its competitor will match that price. An equilibrium identifies a set of outputs where neither firm can improve its own revenue by changing its own price, assuming its competitor has the option of matching or not matching that price. Assume that an intersection point exists that cannot be an equilibrium. Without loss of generality, let firm 1 be the one that changes its prices from (p_1, p_2) to (p_1^*, p_2^*), which is the best set of prices for firm 1 under these conditions. For this assumption to be true, firm 1 must benefit by changing prices if firm 2 does not follow, firm 1 must not benefit by changing prices if firm 2 does follow, and firm 2 would not follow. We will now show that in each case, one of these conditions is violated. There are three cases:

(i) $p_1^* \geq p_1, p_2^* \geq p_2$: If it is possible for firm 1 to increase prices and still use all its capacity, then by proposition 3, it is also possible for firm 2 to profit by increasing prices. If this is not possible, then by assumption $3f$, firm 1 would not choose to do this.

(ii) $p_1^* \leq p_1, p_2^* \leq p_2$: If the price drop does not motivate any new customers to want to subscribe, then firm 1 will decrease its revenue. If this price change leaves firm 2 at less than full capacity, firm 2 would choose to decrease its prices too by assumption $3f$. If the price change does bring in new potential customers and firm 2 cannot take any of them, then firm 1 would profit from increasing prices again. Thus, this violates the assumption that (p_1^*, p_2^*) is the best choice of prices.

(iii) $p_i^* \leq p_i, p_j^* \geq p_j, i \neq j$: Class i customers would switch to firm 1, and class j customers would switch to firm 2. If this change would leave firm 1 at less than full capacity, firm 1 would not do it by assumption $3f$. If this change left firm 2 at less than full capacity, then firm 2 would follow suit. If firm 1 carries nothing but class i and firm 2 carries some class i, then firm 1 has selected a price p_i that is too low. The only other possibility is that firm 2 carries nothing but class j, and then firm 2 would profit by changing its prices to follow firm 1, thereby increasing p_{2j}. *Q.E.D.*

Proposition 5: Under assumptions 1, 2 and 3, if R_1 has only one local maximum over x_{11} for every x_{21}, and the admissible regions $a_1(x_{11})$ and $a_2(x_{21})$ are differentiable, then firm 1's reaction function is continuous in x_{11}.

Proof: As discussed in Section 2.2, if R_1 has only one local maximum over x_{11} for every x_{21}, the reaction function is either defined by $\frac{dR_1}{dx_{11}} = 0$ or the maximum occurs when $x_{11} = 0$. In the latter case, R_1 is a decreasing function, and the optimal reaction is always $x_{11} = 0$. Otherwise, it is the solution to:

$$\frac{dR_1}{dx_{11}} = B_1 x_{11} \frac{df_1(s_1)}{ds_1} + B_1 f_1(s_1) + B_2 a_1(x_{11}) \frac{df_2(s_2)}{ds_2} \frac{da_1(x_{11})}{dx_{11}} + B_2 f_2(s_2) \frac{da_1(x_{i1})}{dx_{11}} = 0$$

$a_1(x_{11})$ and x_{11} are continuous by definition. $\frac{da_1(x_{11})}{dx_{11}}$ is continuous by assumption. By assumptions $1(d)$ and $1(e)$, $\frac{df_1(x_1)}{dx_1}$ and $\frac{df_2(x_2)}{dx_2}$ must also be continuous. Hence $\frac{dR_1}{dx_{11}}$ is continuous. If R_1 has only one maximum over x_{11} for every x_{21}, the equation $\frac{dR_1}{dx_{11}} = 0$ defines the reaction function for firm 1. Since $\frac{dR_1}{dx_{11}}$ is continuous, the reaction function is also continuous. *Q.E.D.*

<u>*Proposition 6:*</u> Under assumptions 1, 2 and 3, and if, for $0 < x_{11} < x_{11}^{max}$ and $0 < x_{21} < x_{21}^{max}$:
(i) $\frac{d^2 f_1(x_1)}{dx_1^2}, \frac{d^2 f_2(x_2)}{dx_2^2} \leq 0$,
(ii) $\frac{d^2 a_1(x_{11})}{dx_{11}^2} \leq 0$, and
(iii) $\frac{dR_{11}}{dx_{11}} \geq 0$ or $\frac{dR_{12}}{dx_{12}} \geq 0$ over all x_{11},
then R_1 has only one local maximum over x_{11} for every x_{21}.

Proof: Since $a_1(x_{11})$, p_1 and p_2 are continuous, $R_1(x_{11})$ is continuous in x_{11}. Therefore if there are two local maxima, there must be a local minimum between the two. At any local minimum, the slope is zero and the second derivative is positive. The first derivative is

$$\frac{dR_1}{dx_{11}} = B_1 \frac{dR_{11}}{dx_{11}} + B_2 \frac{dR_{12}}{dx_{12}} \frac{da_{11}}{dx_{11}}$$

Since B_1 and $B_2 > 0$, and $\frac{da_{11}}{dx_{11}} \leq 0$, the first derivative can only equal 0 if $\frac{dR_{11}}{dx_{11}}$ and $\frac{dR_{12}}{dx_{12}}$ are both positive, both negative, or both 0. By condition (iii), this means that $\frac{dR_{12}}{dx_{12}} \geq 0$. The second derivative of revenue is:

$$\begin{aligned} \frac{d^2 R_1}{dx_{11}^2} &= 2B_1 \frac{df_1(x_1)}{dx_{11}} + B_1 x_{11} \frac{d^2 f_1(x_1)}{dx_{11}^2} + 2B_2 \frac{df_2(x_2)}{dx_{12}} \left(\frac{da_1(x_{11})}{dx_{11}}\right)^2 + \\ & B_2 a_1(x_{11}) \frac{d^2 f_2(x_2)}{dx_{12}^2} \left(\frac{da_1(x_{11})}{dx_{11}}\right)^2 + B_2 \frac{dR_{12}}{x_{12}} \frac{d^2 a_1}{dx_{11}^2} \end{aligned}$$

where $x_1 = x_{11} + x_{21}$ and $x_2 = x_{12} + x_{22}$. Terms 1 and 3 of $\frac{d^2 R_1}{dx_{11}^2}$ are always negative. Condition (i) implies that terms 2 and 4 are negative. Given condition (ii), if $\frac{dR_{12}}{x_{12}} > 0$, then term 5 is negative. Hence the second derivative is always negative when the first derivative is 0, and there cannot be two local maxima. *Q.E.D.*

<u>*Proposition 7:*</u> Under assumptions 1, 2 and 3, if the reaction function is continuous and the price functions are concave, then the reaction function is monotonically decreasing.

Proof: Let $r_1(x_{21})$ be firm 1's reaction function.

$$\begin{aligned} \frac{dr_1(x_{21})}{dx_{21}} &= -(B_1 \frac{df_1(x_{11} + x_{21})}{dx_{21}} + B_2 \frac{df_2(x_{12} + x_{22})}{dx_{12}} \frac{da_1 11}{dx_{11}} \frac{dx_{22}}{dx_{21}} + \\ & B_2 a_1(x_{11}) \frac{d^2 f_2(x_{12} + x_{22})}{dx_{12}^2} \frac{da_1(x_{11})}{dx_{11}} \frac{da_2(x_{21})}{dx_{21}} + B_1 x_{11} \frac{d^2 f_1(x_{11} + x_{21})}{dx_{11}^2}) / \frac{d^2 R_1}{dx_{11}^2} \end{aligned}$$

Since the denominator is negative, the sign of $\frac{dr_1(x_{21})}{dx_{21}}$ is the same as that of the numerator. Terms 1 and 2 of the numerator are negative. The slope of admissible load region is negative for both firms. Thus, terms 3 and 4 are negative if the price functions are concave. Hence, $\frac{dr_1(x_{21})}{dx_{21}} < 0$. *Q.E.D.*

Proposition 8: Under assumptions 1, 2, 3, and if, *either*
(i) the reaction functions of both firms are continuous, *or*
(ii) the reaction function for one firm is continuous and both firms have either a monotonically increasing or a monotonically decreasing reaction function,
then at least one equilibrium exists.

Proof: Firm 1's reaction function $r_1(x_{21})$ is within the range $[0, x_{11}^{max}]$ for all x_{21} within the range $[0, x_{21}^{max}]$. Similarly, firm 2's reaction function $r_2(x_{11})$ is within the range $[0, x_{21}^{max}]$ for all x_{11} within the range $[0, x_{11}^{max}]$. $r_1(0)$ is greater than or equal to any point in the range of $r_2(x_{11})$, and $r_1(x_{21}^{max})$ is less than or equal to any point in the range of $r_2(x_{11})$. If $r_1(x_{21})$ and $r_2(x_{11})$ are both continuous, then at some point between $r_1(0)$ and $r_1(x_{21}^{max})$, the two reaction functions must intersect. The same is also true if the reaction function for one firm is continuous and both firms have either a monotonically increasing or a monotonically decreasing reaction function. *Q.E.D.*

Proposition 9: Under assumptions 1, 2 and 3, if the firms have identical admissible load regions, and the reaction functions are continuous, then an equal-output equilibrium always exists.

Proof: Since the admissible load regions are identical, the coefficient of x_{11} in $\frac{dR_1}{dx_{11}} = 0$ is the same as the coefficient of x_{21} in $\frac{dR_2}{dx_{21}} = 0$. Hence, the reaction functions for the two firms are symmetric. Therefore, if $(x_{21} = x^*,\ x_{11} = x^*)$ is on firm 1's reaction function, it must be on firm 2's reaction function also. Since the reaction functions are continuous, they must cross the $x_{11} = x_{21}$ line at least once. Hence there will be at least one equal-output equilibrium. *Q.E.D.*

Proposition 10: Under assumptions 1, 2 and 3, if the firms have identical admissible load regions and the reaction functions are continuous, unequal-output equilibria can occur only in symmetric pairs.

Proof: Since the admissible load regions are identical, the reaction functions of the two firms are symmetric. Therefore if $(x_{11} = x_1^a,\ x_{21} = x_1^b)$ is an equilibrium, $(x_{11} = x_1^b,\ x_{21} = x_1^a)$ must also be an equilibrium. Thus, the unequal-output equilibria occur in pairs. *Q.E.D.*

Proposition 11: Under assumptions 1, 2 and 3, if the firms have identical admissible load regions, and the boundary of the admissible load region $a(x_{i1})$ is linear or concave, then unequal-output equilibria do not exist.

Proof: We will show, by contradiction, that unequal-output equilibria are not possible when firms have identical linear or concave admissible load regions. Let $(x_{11} = x_{11}^*, x_{21} = x_{21}^*)$, where $x_{11}^* < x_{21}^*$, be an unequal-output equilibrium. Thus, $\frac{dR_2}{dx_{21}} = 0$ and $\frac{dR_1}{dx_{11}} = 0$ at $(x_{11} = x_{11}^*, x_{21} = x_{21}^*)$.

$$\begin{aligned}\frac{dR_1}{dx_{11}} - \frac{dR_2}{dx_{21}} \quad = \quad & B_1(x_{21}^*\frac{df_1(x_1)}{dx_{21}} - x_{11}^*\frac{df_1(x_1)}{dx_{11}}) + B_2 f_2(x_2)(\frac{da(x_{21}^*)}{dx_{21}} - \frac{da(x_{11}^*)}{dx_{11}}) + \\ & B_2(a(x_{21}^*)\frac{df_2(x_2)}{dx_{21}} - a(x_{11}^*)\frac{df_2(x_2)}{dx_{11}}) = 0\end{aligned}$$

where $x_1 = x_{11} + x_{21}$ and $x_2 = x_{12} + x_{22}$. Since $\frac{df_1(x_{11}+x_{21})}{dx_{11}} < 0, \frac{df_1(x_{11}+x_{21})}{dx_{21}} < 0$, and $x_{11}^* < x_{21}^*$, $(x_{21}\frac{df_1(x_{11}+x_{21})}{dx_{21}} - x_{11}\frac{df_1(x_{11}+x_{21})}{dx_{11}}) < 0$. Since $\frac{da(x_{i1})}{dx_{i1}} < 0$, $a(x_{21}^*) < a(x_{11}^*)$. Since $\frac{df_2(x_{12}+x_{22})}{dx_{21}} > 0$ and $\frac{df_2(x_{12}+x_{22})}{dx_{11}} > 0$, $(a(x_{21}^*)\frac{df_2(x_{12}+x_{22})}{dx_{21}} - a(x_{11}^*)\frac{df_2(x_{12}+x_{22})}{dx_{11}}) < 0$. Since the boundary of the admissible load region, $a(x_{i1})$, is concave and $x_{11} < x_{21}$, $\frac{da(x_{21}^*)}{dx_{21}} < \frac{da(x_{11}^*)}{dx_{11}}$. Hence, $(\frac{dR_2}{dx_{21}} - \frac{dR_1}{dx_{11}}) \neq 0$. Therefore, $(x_{11} = x_{11}^*, x_{21} = x_{21}^*)$ can not be an equilibrium. *Q.E.D.*

Proposition 12: Under assumptions 1, 2 and 3, if the firms have identical admissible load regions, and *either*
(i) $a(x_{i1})$, the boundary of the admissible load region, is linear or concave, *or*
(ii) $\frac{dR_{12}}{dx_{12}} \leq 0$ for all x_{11},
then zero-output equilibria do not exist.

Proof: Zero-output equilibria are a special type of unequal-output equilibria, and by Proposition 11, unequal-output equilibria can not occur when admissible load regions have linear or concave boundaries. As for condition (ii), let $(x_{11} = 0, x_{21} = x_{21}^*)$ be a zero-output equilibrium. $\frac{dR_1(x_{11}=0)}{dx_{11}} \leq 0$ is a necessary condition for equilibrium. $\frac{dR_1(x_{11}=0)}{dx_{11}} = B_1 f_1(x_{21}^*) + B_2(f_2(x_{12} + x_{22}) + a(x_{11} = 0)\frac{df_2(x_{12}+x_{22})}{dx_{12}})\frac{da(x_{11}=0)}{dx_{11}}$. If $\frac{dR_{12}}{dx_{12}}(= B_2 f_2(x_{12} + x_{22}) + B_2 a(x_{11})\frac{df_2(x_{12}+x_{22})}{dx_{12}}) \leq 0$, $\frac{dR_1(x_{11}=0)}{dx_{11}}$ will be positive. Hence, $(x_{11} = 0, x_{21} = x_{21}^*)$ can not be an equilibrium. *Q.E.D.*

References

[1] Daugherty, A. F., 1988, editor, Cournot oligopoly - characterization and applications (Cambridge University Press, Cambridge).

[2] Davidson, C. and Deneckere, R., 1986, Long-run competition in capacity, short-run competition in price, and the Cournot model, Rand Journal of Economics, Vol. 17, No. 3, 404-415.

[3] Demers, A., Keshav, S., and Shenker, S., 1989, Analysis and simulation of a fair queueing algorithm, in Proceedings of ACM Sigcomm, 1-12.

[4] Friedman, J., 1983, Oligopoly Theory, (Cambridge University Press, Cambridge).

[5] Golestani, S. J., 1991, A framing strategy for congestion management, IEEE Journal on Selected Areas in Communications, Vol. 9, No. 7, 1064-77.

[6] Herk, L. F., 1993, Consumer choice and Cournot behavior in capacity-constrained duopoly, Rand Journal of Economics, Vol. 24, No. 3, 399-417.

[7] Hyman, J., Lazar, A. A., and Pacifici, G., 1991, Real-time scheduling with quality of service constraints, IEEE Journal on Selected Areas in Communications, Vol. 9, No. 7, 1052-63.

[8] Kalmanek, C. R., Kanakia, H., and Keshav, S., 1990, Rate controlled servers for very high-speed networks, in Proceedings of IEEE Globecom, 12-20.

[9] Kreps, D. M. and Scheinkman, J. A., 1983, Quantity precommitment and Bertrand competition yield Cournot outcomes, Bell Journal of Economics, Vol. 14, No. 2, 326-37.

[10] Lee, H. W., and Mark, J. W., 1995, Capacity allocation in statistical multiplexing of ATM sources, IEEE/ACM Transactions on Networking, Vol. 3, No. 2, 139-151.

[11] Lynn, M. A., and Peha, J. M., 1997, Priority token bank scheduling in a network of queues, Proceedings of IEEE International Conference on Communications (ICC), pp. 1387-91.

[12] Maglaris, B., Anastassiou, D., Sen, P., Karlsson, G., and Robbins, J. D., 1988, Performance models of statistical multiplexing in packet video communications, IEEE Transactions on Communications, Vol. 36, No. 7, 834-44.

[13] Novshek, W., 1985, On the Existence of Cournot Equilibrium, Review of Economic Studies 52, 85-98.

[14] Parekh, A. K. and Gallager, R. G., 1994, A generalized processor sharing approach to flow control in integrated services networks: the multiple-node case, IEEE/ACM Transactions on Networking, Vol. 2, No. 2, 137-50.

[15] Peha, J. M., 1997, Dynamic pricing as congestion control in ATM networks, accepted to appear in Proceedings of IEEE Globecom.

[16] Peha, J. M., 1996b, Simulating ATM integrated-services networks, Proceedings of 29th Annual IEEE/ACM/SCS Simulation Symposium, 162-171.

[17] Peha, J. M. and Tobagi, F. A., 1996a, Cost-based scheduling and dropping algorithms to support integrated services, IEEE Transactions on Communications, Vol. 44, No. 2, 192-202.

[18] Peha, J. M., 1995, Heterogeneous-criteria scheduling: minimizing weighted number of tardy jobs and weighted completion time, Computers and Operations Research, vol. 22, no. 10, 1089-1100.

[19] Peha, J. M., 1993, The priority token bank: integrated scheduling and admission control for an integrated-services network, Proceedings of IEEE International Conference on Communications (ICC), 345-51.

[20] Sidhu, I. and Jordan, S., 1995, Multiplexing gains in bit stream multiplexors, IEEE/ACM Transactions on Networking, Vol. 3, No. 6, 785-97.

[21] Tirole, J. and Fudenberg, D., 1991, Game theory (MIT Press, Cambridge) 23-29.

[22] Wang, Q., Peha, J. M., and Sirbu, M., 1997, Optimal pricing for integrated-services networks with guaranteed quality of service, Internet Economics, J. Bailey and L. McKnight eds., MIT Press, 353-76.

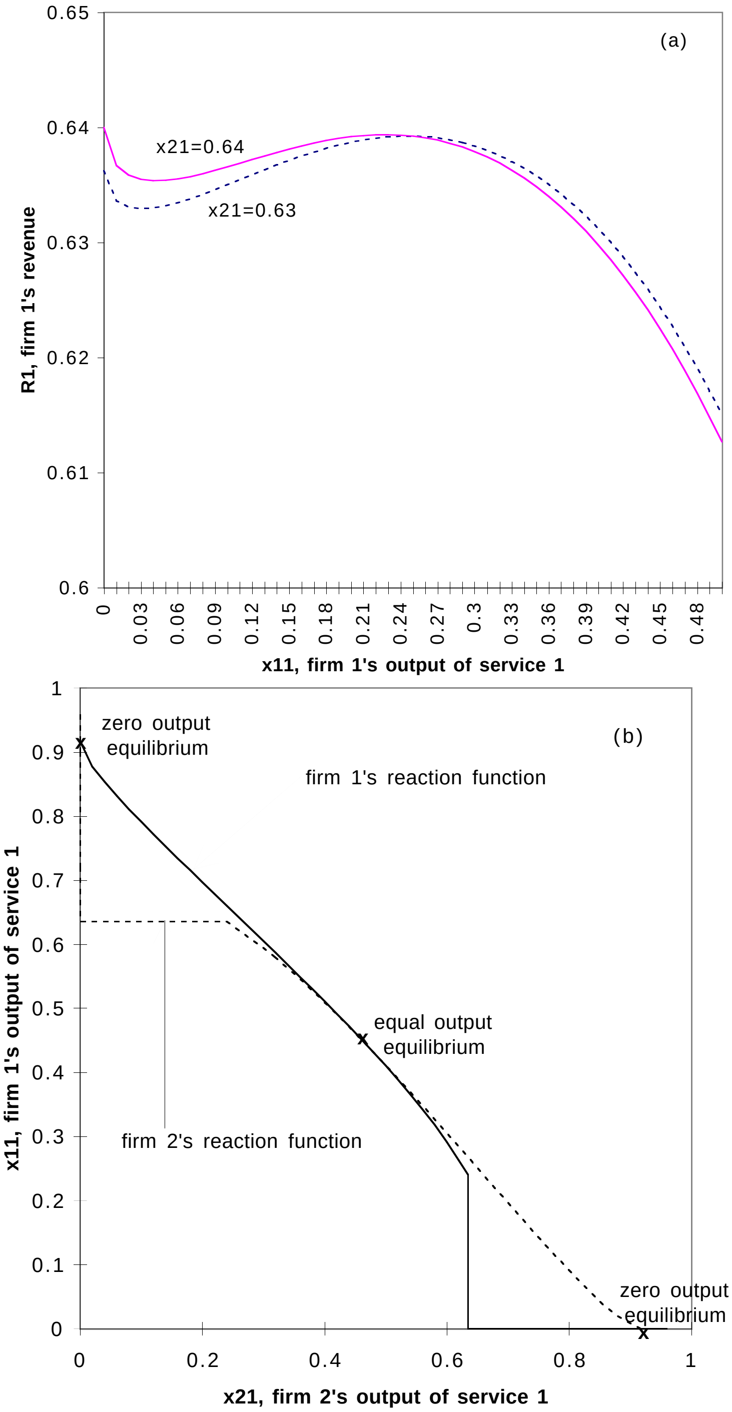

Figure 1: (a) Two local maxima, (b) Discontinuous reaction function, Multiple and Zero-output equilibria; $f_i(x_i) = 1 - (\frac{x_i}{2})^2, a_j(x_{j1}) = 1 - (x_{j1})^{0.5}$.

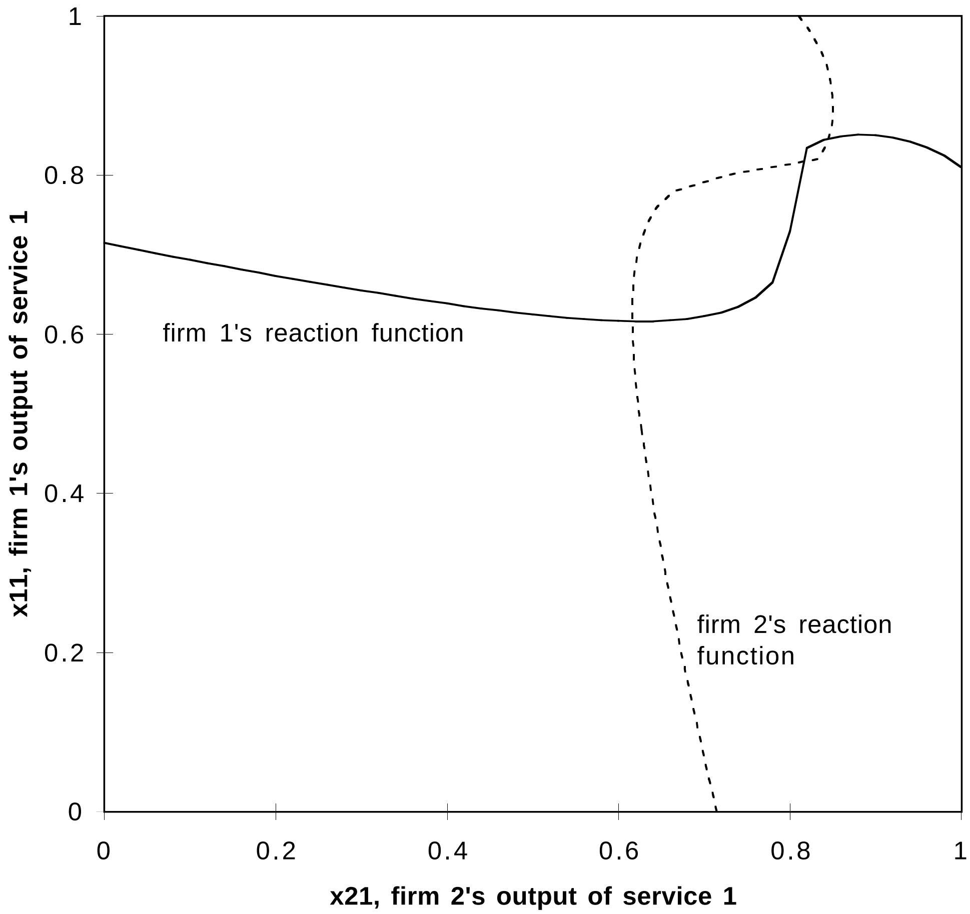

Figure 2: Non-monotonic reaction function. Multiple equal-output equilibria: $f_i(x_i) = 3 - 4(\frac{x_i}{1.5}) + (\frac{x_i}{1.5})^2, a_j(x_{j1}) = 1 - (x_{j1})^7$.

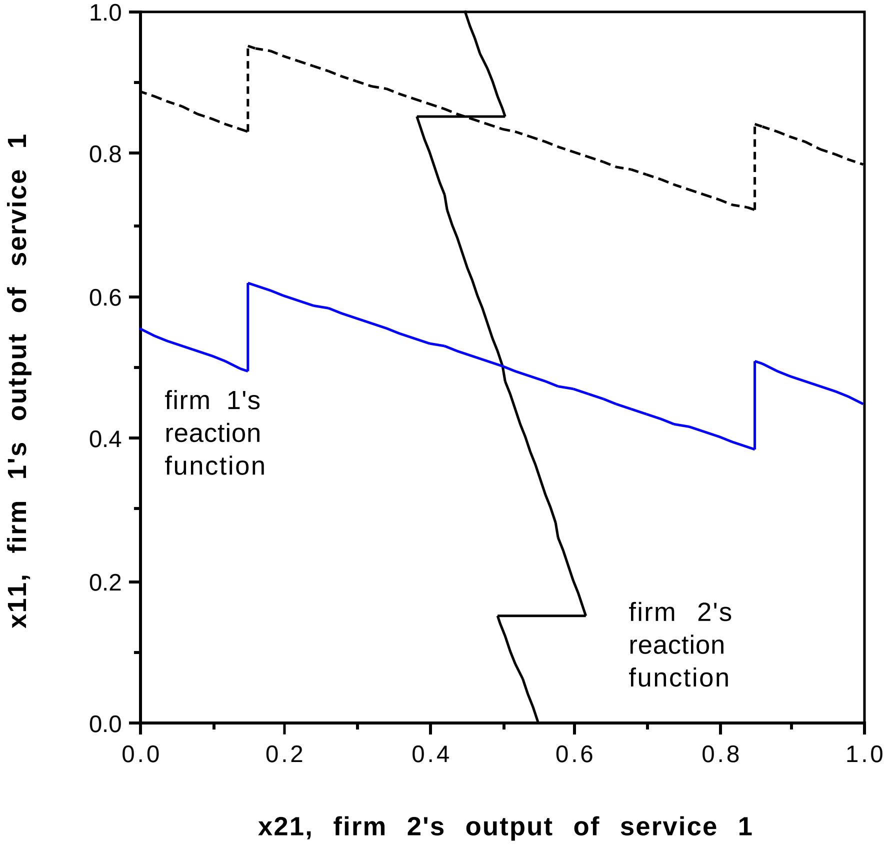

Figure 3: Discontinuous reaction functions. Solid lines: $f_i(x_i) = max(x_i^2 - 4x_i + 3.5,\ 0), a_j(x_{j1}) = 1 - x_{j1}$. Dashed line: firm 1's reaction function + constant.

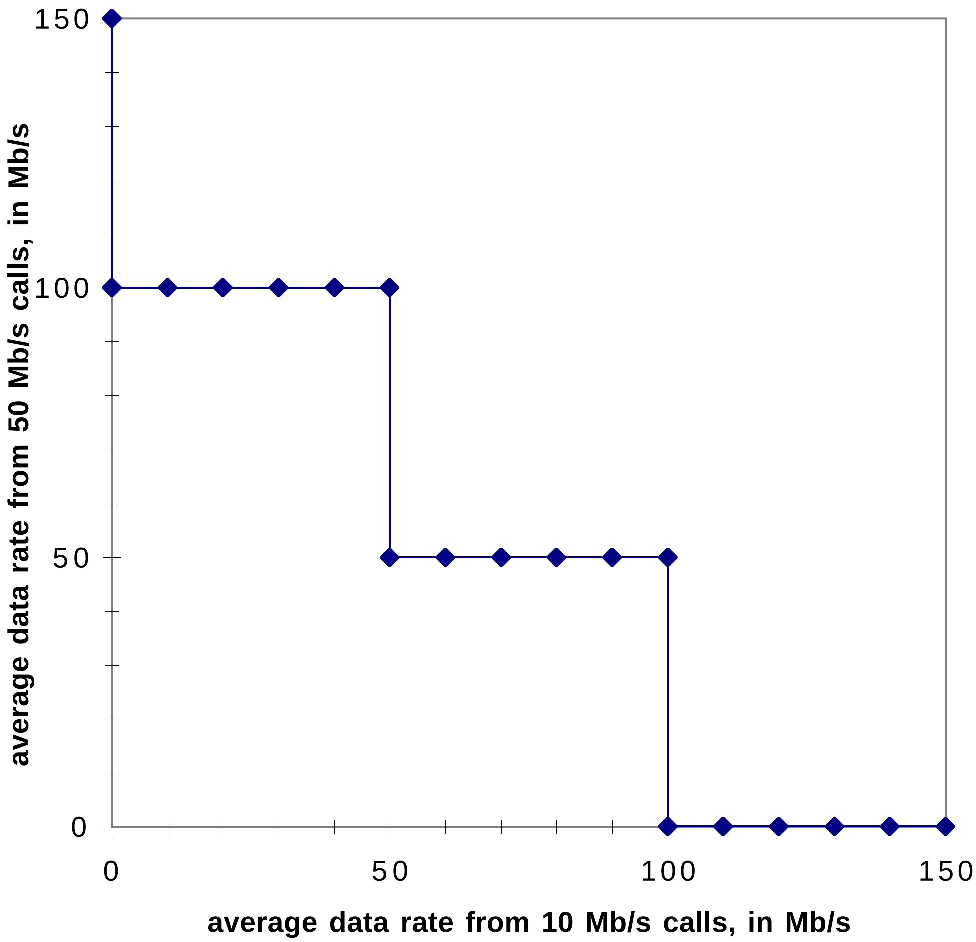

Figure 4: Schedulable Region with constant-bit-rate 50 Mb/s and 10 Mb/s calls in a 150 Mb/s channel.

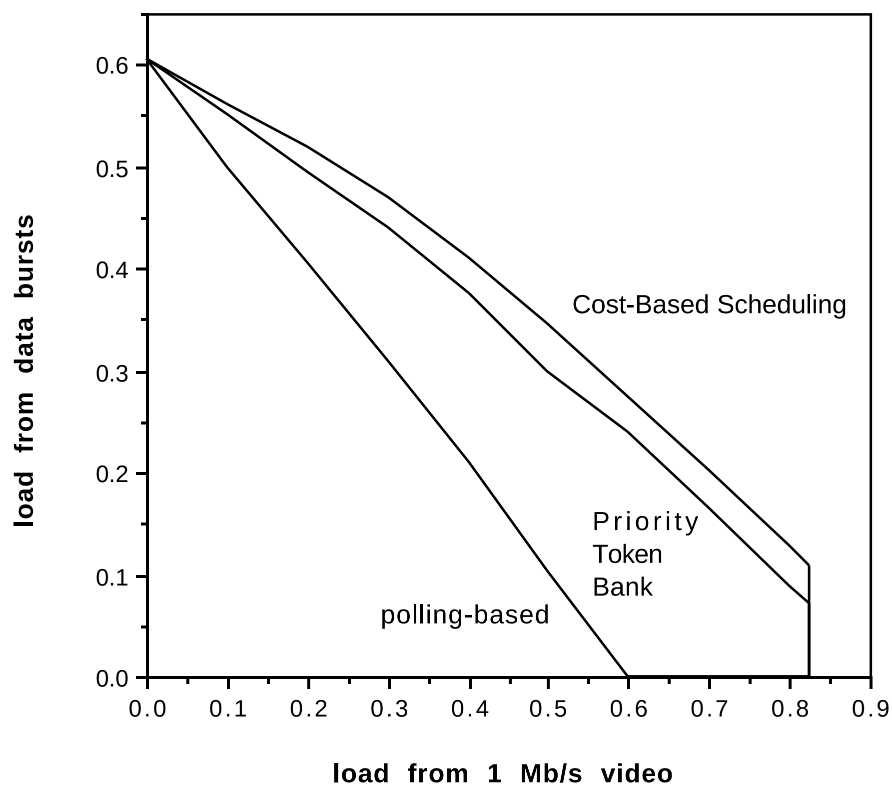

Figure 5: Schedulable Regions for video and data bursts in a 150 Mb/s channel. 1 Mb/s VBR video and Poisson arrival of data bursts with lengths that are exponentially distributed with mean 500 kb.

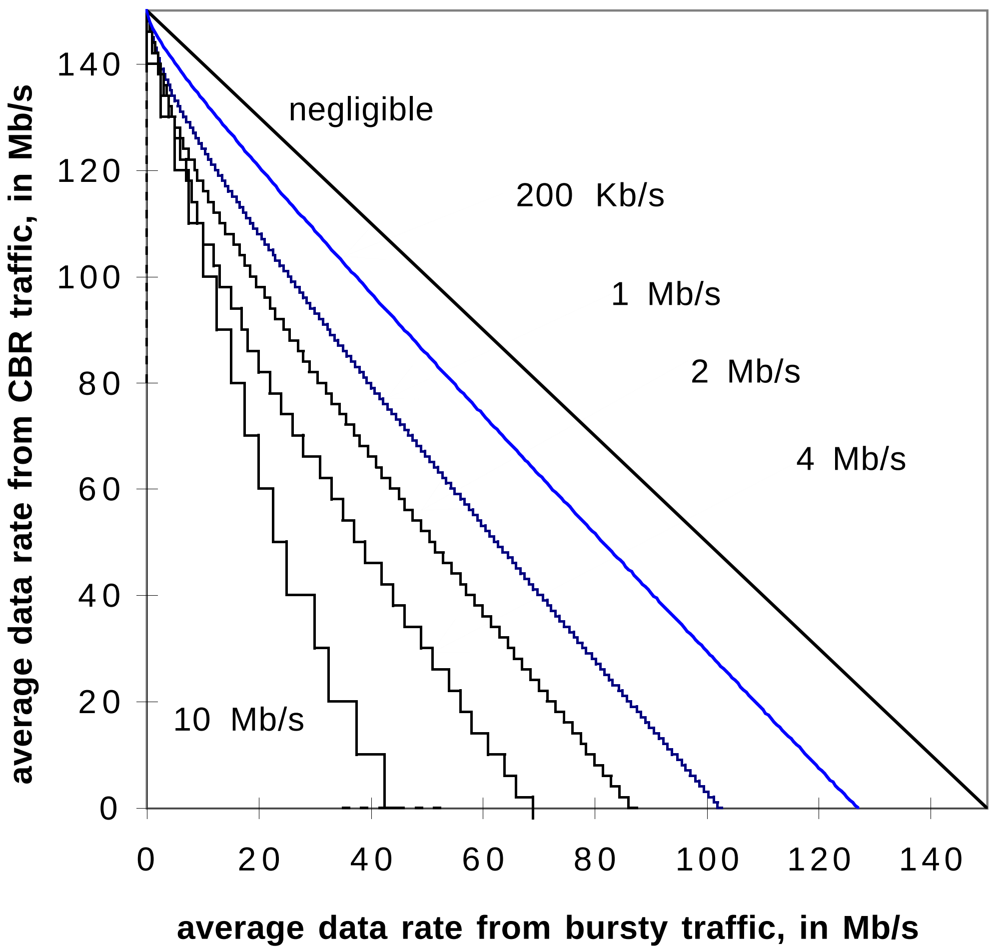

Figure 6: Schedulable Regions with VBR and CBR calls in a 150 Mb/s channel, with peak data rate of negligible to 10 Mb/s from the VBR source, which is modeled as an on-off source with exponentially distributed on and off periods with 100 ms and 300 ms means, respectively.

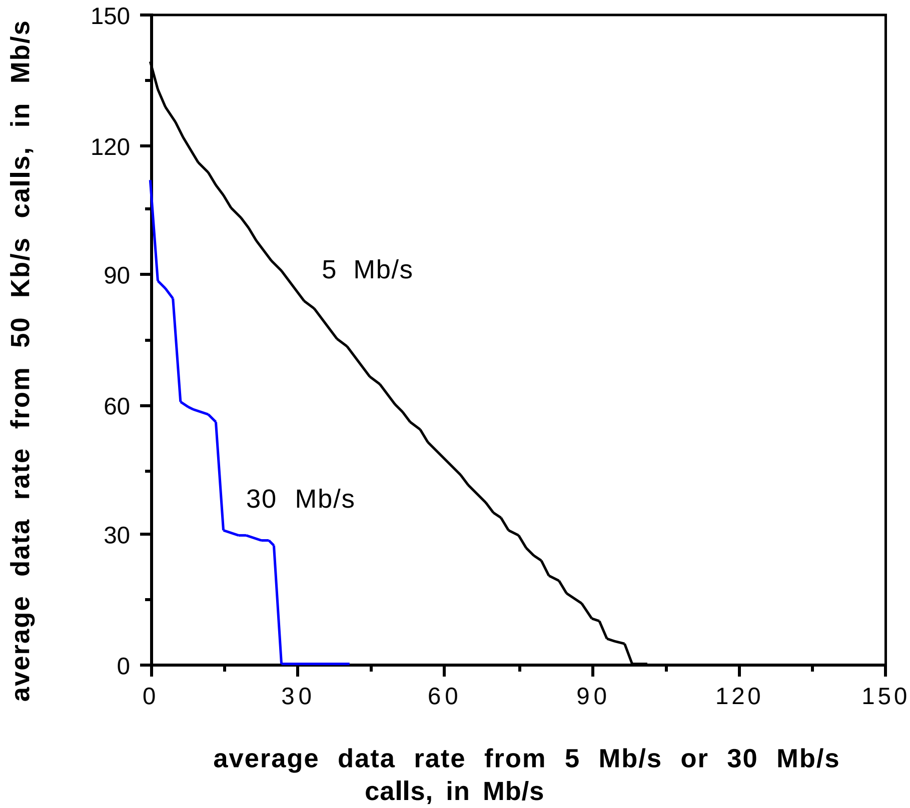

Figure 7: Admissible load Region with circuit-switched 5Mb/s or 30 Mb/s calls and 50Kb/s calls in a 150 Mb/s channel (Poisson call arrival process, exponential call duration, .1%blocking probability for both services).

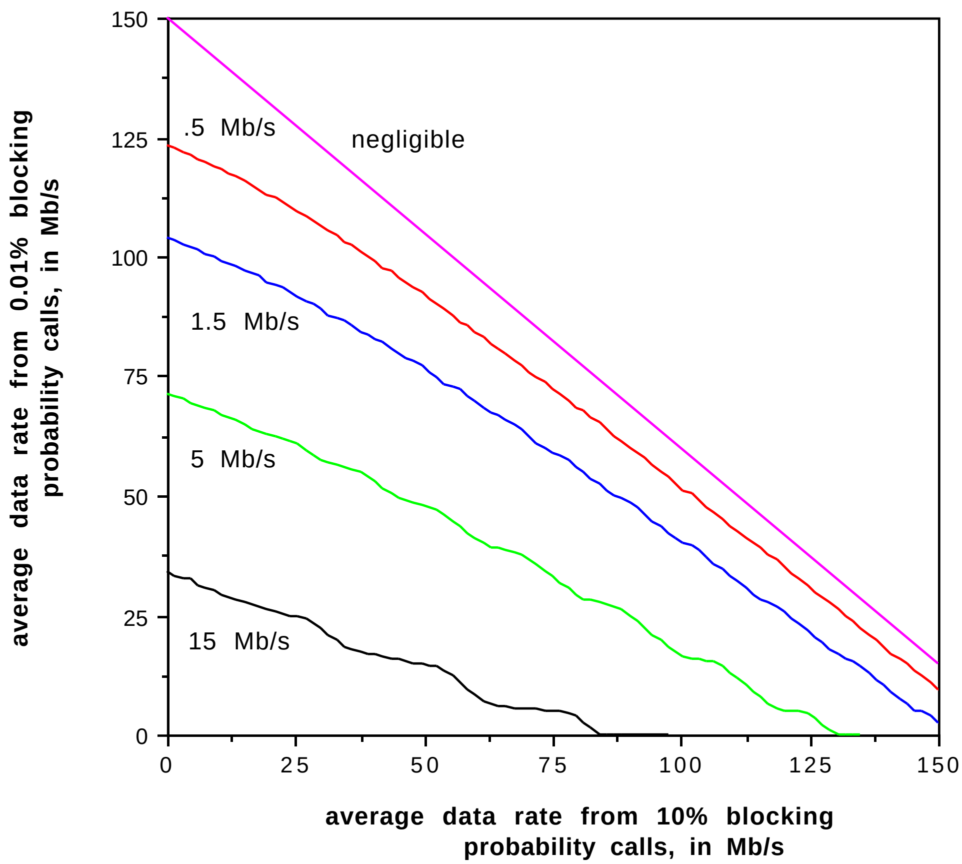

Figure 8: Admissible Load Regions with CBR services that tolerate .01% and 10% blocking probability, with data rates from negligible to 15 Mb/s (Poisson call arrival process, exponential call duration).