ON PRICING STRATEGIES FOR STOCHASTIC CAPACITY NETWORKS

Carey Williamson   Hongxia Sun
Department of Computer Science, University of Calgary
Calgary, AB, Canada T2N 1N4
Email: {carey,sunh}@cpsc.ucalgary.ca

Abstract

This paper studies pricing strategies for networks with stochastic capacity variation. Call-level simulation is used to compare the profit generated with four different pricing models, and with different policies for the management of call dropping episodes in the network. Both user-oriented and network-oriented performance metrics are considered. The simulation results show that the choice of an optimal pricing strategy depends on the call dropping control policy used in the network, and vice versa. The *NewestArrival* call dropping policy provides surprisingly robust performance for the pricing strategies considered.

Keywords: Network Management and Control, Stochastic Capacity Networks, Pricing, Call Dropping, Simulation

1 Introduction

In conventional network environments, such as wired networks, the physical transmission capacity is a fixed quantity. In such networks, the only stochastic characteristics of interest arise from the offered network traffic. These networks have been well-studied for many years, with the emphasis on topics such as admission control, quality of service, congestion control, and fairness.

In other network environments, including wireless networks, the available network capacity may vary unpredictably with time [1, 2, 4, 6, 11, 12]. This phenomenon can occur in CDMA cellular systems, wireless LANs, and wireless ad hoc networks, where capacity variation arises from the mobility of users and the time-varying characteristics of the wireless propagation environment [1, 4, 6].

The traditional techniques used to deal with capacity variation problems are admission control, resource reservation, or adaptive rate control. For example, in wireless CDMA systems, dynamic power control and rate adaptation are used to reduce the aggregate data rate when capacity problems occur [10]. This approach maintains all active calls, but with degraded service quality.

In this paper, we consider a different approach to the stochastic capacity problem, namely *call dropping*. This approach removes (drops) selected calls from the network when the traffic demands temporarily exceed the available resources in the network, due to a capacity decrease.

In most networks, call dropping is a last resort. Call blocking at the time of call arrival is deemed tolerable, as long as it is low (e.g., at most 2%). In fact, call blocking performance is the key metric used in the capacity planning process for traffic engineering (e.g., Erlang B blocking formula). Dropping a user in the middle of a call (e.g., failed handoff [3]) is deemed unacceptable, since it disrupts users, violates service agreements, and wastes the network resources consumed.

Our work considers a network model in which the physical network capacity has stochastic characteristics. In such a network, call dropping episodes are inevitable, particularly when the network is operating at high utilization when a capacity decrease occurs. A carefully chosen call dropping policy is important to maintain overall system performance, particularly when the traffic demands and the capacity variation have complex statistical behaviours.

The purpose of our paper is to explore pricing models for stochastic capacity networks. Call dropping controls are used to make judicious decisions about which calls are disrupted during the dropping episodes, with the intent of minimizing the number of disrupted calls. The dropping control mechanisms have a noticable impact on call-level performance, in terms of call blocking, call dropping, and network utilization. When coupled with different pricing strategies, additional tradeoffs occur amongst the different performance metrics, particularly between user-oriented quality of service (QoS) metrics (e.g., call blocking, call dropping) and the metrics of interest to network service providers (e.g., utilization, revenue, profit).

Our paper makes three main contributions. First, we show via call-level simulation that call dropping policies can have an important impact on call-level performance in a stochastic capacity network. Second, we propose and evaluate four possible pricing strategies for stochastic capacity networks. Third, we demonstrate the performance tradeoffs that occur between call blocking, call dropping, and profit maximization in stochastic capacity networks.

2 Related Work

There are many papers in the literature discussing variable-capacity networks [5, 8, 10, 11, 12]. However, most of this work [8, 9, 10, 11, 12] focuses on Connection Admission Control (CAC), rather than call dropping controls or network pricing strategies.

Our work draws partly upon ideas from *performability* modeling [3, 7], which takes into account both performance and availability of a system. Traditional performance modeling ignores stochastic variation of a system's shared resources, and can overestimate system performance. Meanwhile, availability analysis tends to be conservative, since performance considerations are not taken into account. Performability modeling, however, can provide a complete picture for system performance analysis.

Prior simulation work by Sun and Williamson [13] studies the call-level performance of dropping policies in stochastic capacity networks. However, only user-oriented QoS metrics were considered. The recommended control policies may not be optimal if other optimization criteria, such as network-oriented metrics, are used. Identifying these tradeoffs is the main contribution of our paper.

3 System Model

We model a stochastic capacity network. The network has an overall average capacity for carrying C simultaneously active calls, but the capacity varies randomly with time. The network capacity always has a non-negative integer value, but the capacity changes can occur at arbitrary points in continuous time.

Our model specifies both the timing characteristics of the stochastic capacity process, as well as the distribution for the capacity value. We focus primarily on the Normal distribution, for which it is easy to control both the mean and the variance of the network capacity. Capacity values are drawn from this distribution as independent and identically distributed (iid) samples.

We model a generic call-level workload, suitable for an arbitrary network carrying either voice or data services. New calls arrive according to a specified arrival process, such as Poisson. Each call has a specified holding time, drawn from a specified distribution (e.g., Exponential). Each call requires one unit of network capacity for the duration of the call.

The network uses a simple Greedy CAC algorithm in which a call is admitted into the network if and only if adequate capacity exists for it at the time of call arrival. There is no future lookahead in the CAC mechanism.

3.1 Call Dropping Policies

Call dropping occurs in a stochastic capacity network when the aggregate traffic load currently in the network temporarily exceeds the available network capacity. This phenomenon happens if the network is full or nearly full when a capacity decrease occurs. When this unfortunate network condition occurs, the network must expunge one or more *victim* calls in order to conform to the new capacity constraints. We call such an event a *call dropping episode*.

Choosing which call(s) to drop in a dropping episode is determined by a call dropping policy. These policies impact both user-level and network-level performance.

We consider 7 simple call dropping policies, as in [13]. *Random* chooses uniformly at random amongst the active calls in the network whenever a victim call must be dropped. *NewestArrival* removes the youngest call, and *OldestArrival* removes the oldest call from the network. *EarliestDeparture* removes the active call that is scheduled to complete next, and *LatestDeparture* drops the active call whose departure time is furthest in the future. *ShortestDuration* drops the call with the shortest holding time specified at the time of its arrival, and *LongestDuration* drops the active call with the longest original holding time.

3.2 Pricing Model

Our network revenue model operates as follows. The network provider receives payment from a user for each *Successful* call that is admitted to the network and completed without disruption. This income is referred to as *Revenue*. Conversely, the network provider pays to the user a penalty fee for each *Dropped* call: a call that is initially admitted to the network, but is prematurely disrupted and dropped from the network prior to completion. Note that there is no revenue generated from Dropped calls; only the outflow of penalty payment occurs. These payments are referred to as *Expenses*. Finally, we assume that *Blocked* calls, which are rejected from the network at the time of call arrival, are revenue-neutral. That is, no payment occurs at all, since the user receives no service, and the network does not expend any bandwidth resources on the call.

The primary performance metric of interest is *Profit*, which is defined as $Profit = Revenue - Expenses$. The network provider is interested in maximizing profit, for a given call workload and network capacity conditions. The call dropping controls used in the network influence the number of calls that are *Successful*, *Blocked*, or *Dropped*, while the pricing strategy determines the monetary value associated with each call. We express all revenue, expenses, and profit values in arbitrary monetary units.

We consider two main categories of pricing strategies, namely *Prix Fixé* (fixed price) and *Per Unit* (usage-based) pricing. The fixed price approach sets a universal price for all calls (e.g., 1 monetary unit each), regardless of the call duration. In usage-based pricing, the revenue associated with a *Successful* call is directly proportional to the call duration. Similarly, the penalty associated with a *Dropped* call is directly proportional to the time that the call spent in the network prior to dropping.

Within each of these categories, we consider two variants, namely *single-class* and *multi-class* pricing. The multi-class strategies use three different price levels: a low price for short duration calls (e.g., less than 10 sec), a moderate price for medium duration calls, and a high price for long duration calls (e.g., longer than 30 sec). In our experiments, about one-third of the generated calls are in each category (i.e., 27% short, 35% medium, and 38% long).

The four resulting pricing strategies are called *Prix Fixe 1* (fixed price, single-class), *Prix Fixe 3* (fixed price, three classes), *Per Unit 1* (usage-based, single-class), and *Per Unit 3* (usage-based, three classes) in the paper.

4 Experimental Methodology

Our work is carried out using call-level simulation. The two inputs provided to the simulation are a call workload file and a network capacity file.

The workload files are generated using the call workload parameters indicated in Table 1. We use workload files with 100,000 calls. We consider this trace length adequate to highlight performance differences among the pricing strategies and call dropping policies evaluated. Steady state is reached after about 10,000 calls. The additional run length used enables greater statistical confidence in the results obtained.

Table 1. Call-Level Workload Parameters

Parameter		Levels
Stochastic Traffic	Arrival Process	Poisson
	Holding Time	Exponential
Call Arrival Rate (calls/sec)		0.1 . . . 1.0 . . . 6.0
Mean Call Holding Time (sec)		30

The capacity files are generated using the models and parameters indicated in Table 2. We use capacity files with up to 10,000 capacity change events. In some simulations, only the initial portion of the capacity file is needed, depending on the frequency of capacity changes.

Table 2. Network Capacity Parameter Settings

Parameter		Levels
Mean Time between Capacity Changes (sec)		10, 15, 30, 60, 120
Stochastic Capacity	Change Timing	Deterministic
	Change Value	Normal
Capacity Value (calls)	Mean	40
	Standard Deviation	5

The different call dropping and pricing strategies are modeled within the simulator. We provide each policy with the same input files, so that each policy handles the same traffic demands under the same network conditions. Differences observed in call-level performance reflect differences in the call dropping and pricing strategies used.

Table 3 shows the factors and levels used in our simulation experiments. We explore the impacts of different call dropping policies, as well as a variety of pricing strategies.

The performance metrics of interest fall into two categories. The user-oriented performance metrics are the call blocking probability and the call dropping probability. The network-oriented metrics are revenue, expenses, and profit.

Table 3. Factors and Levels in Call-Level Simulations

Factor		Levels
Call Dropping Policy		Random, NewestArrival, OldestArrival, EarliestDeparture, LatestDeparture, ShortestDuration, LongestDuration
Pricing Strategy	Pricing Model	*Prix Fixe*, *Per Unit*
	Number of Classes	1, 3

5 Simulation Results

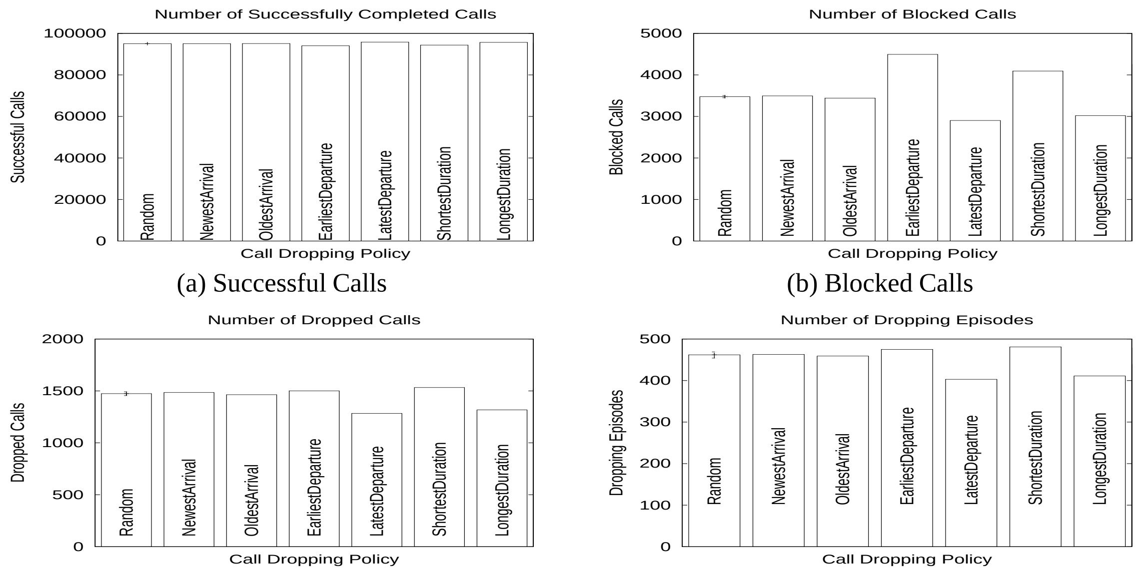
5.1 Overview

Figure 1 presents the call-level performance results from a representative simulation experiment with 100,000 calls. In this experiment, the call arrival process is Poisson, and the call holding times are exponentially-distributed with a mean of 30 seconds. The average offered load is 30 Erlangs. During the simulation, the network capacity varies stochastically, with a random capacity change every 10 seconds. The capacity (in calls) is drawn from a Normal distribution with a mean of 40 and a standard deviation of 5. We chose these parameter values as an approximate model for a typical commercial wireless CDMA system.

There are four graphs in Figure 1. Figure 1(a) shows the number of successful calls for each call dropping policy considered. Figure 1(b) presents the call blocking results, showing the number of offered calls rejected from the network at the time of their arrival. The average blocking rate ranges from 2.9% to 4.5%, depending on the call dropping policy used. Figure 1(c) presents the call dropping results, showing the number of accepted calls that are subsequently dropped from the network prior to completion. The average call dropping rate is lower than the call blocking rate, ranging from 1.3% to 1.5% of the total calls. Figure 1(d) shows the number of call dropping episodes experienced by each call dropping policy. For this workload, there are about 450 call dropping episodes. On average, about 3.2 calls are dropped per dropping episode.

Figure 1 shows that the call dropping policies all achieve a similar number of successful calls. About 95% of the offered calls are *Successful*, regardless of the call dropping policy used. However, the dropping policies do exhibit differences in blocked calls, dropped calls, and dropping episodes. The error bars shown on the *Random* column show the minimum and maximum values from 30 simulation runs, implying that the performance differences seen among the call dropping policies are statistically significant.

It is known that judicious call dropping improves both blocking and dropping performance in the network [13].



(a) Successful Calls (b) Blocked Calls

(c) Dropped Calls (d) Dropping Episodes

Figure 1. Overview of Simulation Results (100,000 calls; PE30; DN(40,5))

For example, the *LatestDeparture* policy is optimal for user-oriented performance. However, this policy is not necessarily optimal if network-oriented performance metrics, such as utilization, revenue, and profit are considered, as we demonstrate next.

5.2 Pricing Results

Next, we vary the pricing strategies to study the effect on revenue, expenses, and network profit. Figure 2 shows these simulation results, for the same network capacity and call workload model as in the previous section.

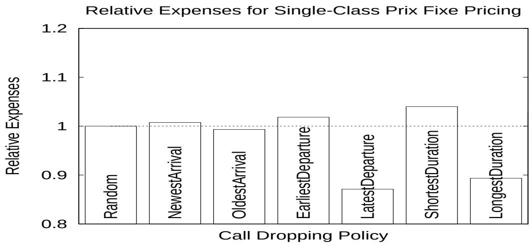
There are four rows of graphs in Figure 2, with each row representing simulation results for a different pricing strategy. In each row, there are two graphs. The graph on the left shows *Expenses* (penalties paid), while the graph on the right shows *Profit* (*Revenue* minus *Expenses*). Revenue is not shown, since it is similar for all dropping policies. For comparison purposes, we normalize all results relative to the values achieved by the *Random* call dropping policy in the same experiment, as shown with the horizontal dashed line. On each graph, there is one pillar shown for each call dropping policy considered. For ease of comparison, the dropping policies are consistently presented in the same relative order in all of the graphs. Note that the three graphs without detailed labels have a different vertical scale than the other five graphs.

Figure 2 illustrates four important observations. First, there are substantial differences among the call dropping policies in terms of the expenses (penalties) that they incur. For example, with the single-class usage-based (*Per Unit 1*) pricing in Figure 2(e), the *OldestArrival* policy incurs excessive penalties, while the *NewestArrival* policy does not. Second, large differences in expenses produce noticeable differences in profits for the call dropping policies. For example, the *NewestArrival* policy outperforms the *OldestArrival* policy with respect to profit in Figure 2(f). Third, the multi-class pricing models tend to alter the differences amongst the call dropping policies for *Prix Fixe*, particularly in terms of expenses (compare Figure 2(a) and Figure 2(c)). This result makes sense since the monetary values are biased across the classes of traffic; some dropping policies exploit this property better than others. Finally, there is no call dropping policy that is universally the best across the set of pricing strategies considered. For *Prix Fixe 1* in Figure 2(b), the best policy is *LatestDeparture*. For this pricing model, profit maximization is actually the same as minimizing dropped calls. For *Per Unit 1* pricing, *LatestDeparture* is inferior to *NewestArrival*, *ShortestDuration*, and two other call dropping policies. The *NewestArrival* policy is surprisingly robust across the range of pricing strategies studied; while not always optimal for profit, this policy is never worse than *Random*, and often better.

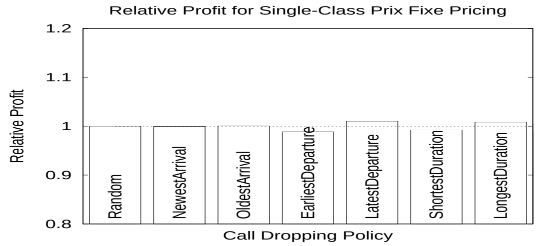
5.3 Effect of Load

The next simulation experiment varies the load level offered to the stochastic capacity network, to study the effect on call-level performance and network profit.

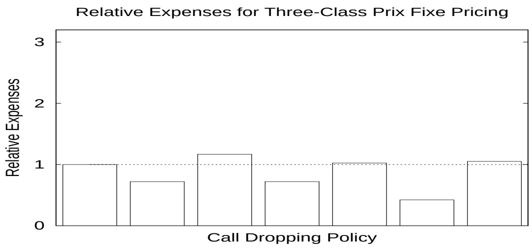
Table 4 summarizes the call-level performance results for this experiment. The results show that call blocking, call dropping, and the number of call dropping episodes all increase with the level of offered load. These results are illustrated using the *Random* call dropping policy as an example; the results for other call dropping policies would differ slightly. The call blocking and call dropping rates in



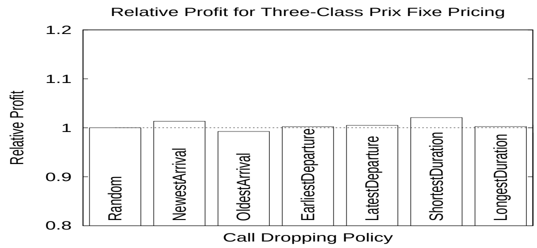
(a) Expenses (Prix Fixe 1)



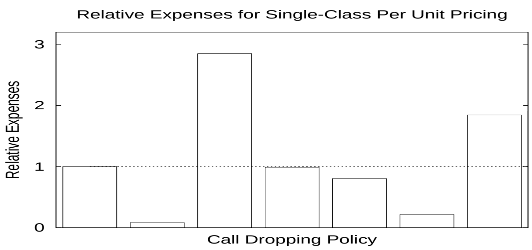
(b) Profit (Prix Fixe 1)



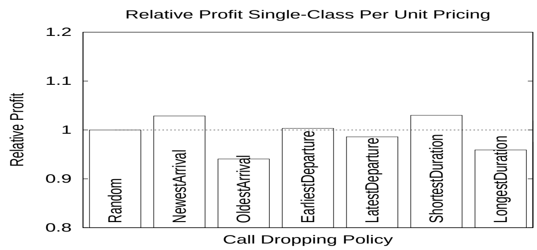
(c) Expenses (Prix Fixe 3)



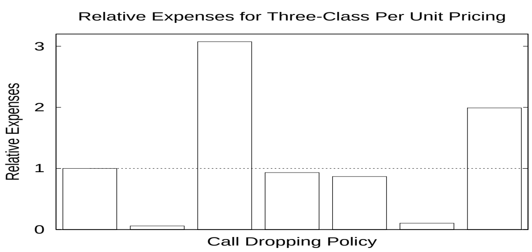
(d) Profit (Prix Fixe 3)



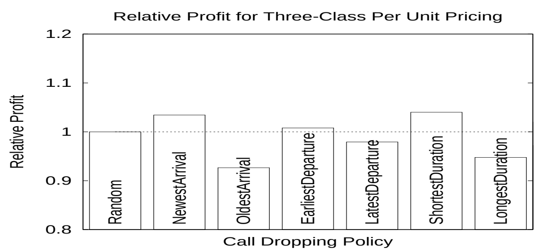
(e) Expenses (Per Unit 1)



(f) Profit (Per Unit 1)

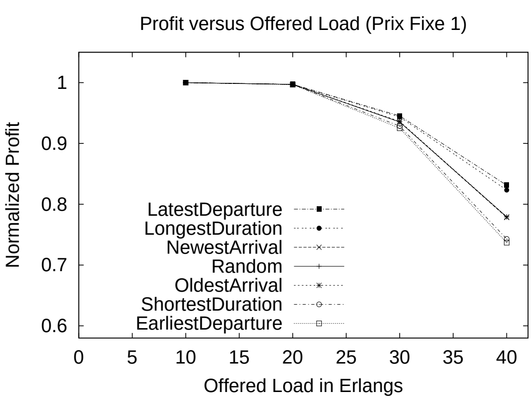


(g) Expenses (Per Unit 3)

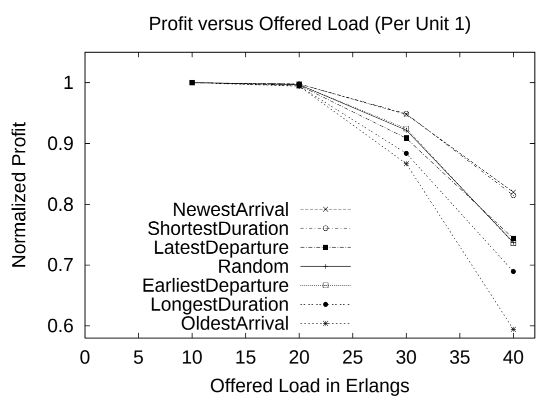


(h) Profit (Per Unit 3)

Figure 2. Simulation Results for Expenses and Profit Relative to Random Call Dropping Policy with Four Different Pricing Models (100,000 calls; PE30; DN(40,5))



(a) Prix Fixe 1



(b) Per Unit 1

Figure 3. Effect of Load Level on Profit (100,000 calls; DN(40,5))

Table 4. Effect of Load Level on Call-Level Performance

Offered Load (Erlangs)	Blocking Rate (%)	Dropping Rate (%)	Dropping Episodes
10	0.005%	0.0%	0
20	0.172%	0.060%	22
30	3.485%	1.496%	465
40	12.187%	4.947%	1,240

Table 4 become excessively high at a load of 40 Erlangs.

Figure 3 shows the effect of offered load on the network profit. These results are presented for two selected pricing strategies, namely *Prix Fixe 1* and *Per Unit 1*. The horizontal axis represents load, which increases from left to right. The vertical axis represents *Normalized Profit*, which is expressed relative to the profit attainable at a load of 10 Erlangs. The 7 lines on each graph represent the profit results from the different call dropping policies simulated. The key on each graph is arranged to match the relative ordering of the lines in the graph.

Figure 3 shows that network profit tends to decrease as network load increases. This result is obvious, since increased load can lead to greater call blocking (less revenue) and the onset of call dropping (more penalties). Both of these effects reduce the network profit.

Figure 3 also shows that the profit differs amongst the 7 call dropping policies, and that the differences depend on the pricing strategy. For example, with *Prix Fixe 1* pricing in Figure 3(a), the best profit occurs with the *LatestDeparture* call dropping policy. For *Per Unit 1* pricing in Figure 3(b), the best policy is *NewestArrival*, followed closely by *ShortestDuration*. *LatestDeparture* performs similarly to *Random*, while *OldestArrival* is the worst.

5.4 Effect of Frequency of Capacity Change

The final simulation experiment varies the relative frequency of the capacity change events in the stochastic capacity network, to understand the effect on call dropping and network profit.

Table 5 summarizes the call dropping performance observed in this experiment, using the *Random* call dropping policy as an example. The first row of this table represents high frequency capacity changes, while the bottom row represents low frequency capacity changes. The number of call dropping episodes decreases when capacity changes are less frequent. Call blocking remains about 3.4% throughout these experiments. Note that a fixed capacity network has no call dropping episodes; only *Successful* and *Blocked* calls are possible.

Figure 4 illustrates the effect on network profit when the frequency of network capacity changes is varied. These results are presented for two selected pricing strategies, namely *Prix Fixe 3* and *Per Unit 3*. The horizontal axis represents the elapsed time between capacity change events, with high frequency changes on the left, and low frequency changes on the right. The vertical axis represents *Normalized Profit*, which is expressed relative to the profit attainable in a static capacity network with a capacity for 40 calls. The 7 lines on each graph represent the different call dropping policies simulated. Again, the key on each graph is arranged to match the relative ordering of the lines.

Table 5. Effect of Frequency of Capacity Changes on Call-Level Performance

Timing (sec)	Blocking Rate (%)	Dropping Rate (%)	Dropping Episodes
10	3.506%	1.465%	460
20	3.343%	0.917%	271
30	3.365%	0.572%	175
60	3.390%	0.319%	93
120	3.229%	0.211%	55

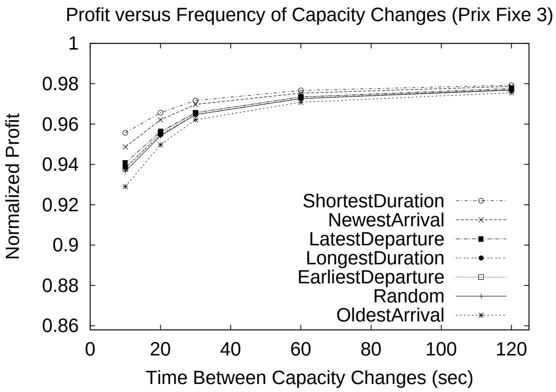
Figure 4 illustrates two main points. First, the achievable profit tends to increase as the time between capacity changes increases. This result makes sense since fewer call dropping episodes occur, and there are fewer penalties paid. All call dropping policies asymptotically approach the same profit value when dropping episodes are rare. Second, the relative ordering of call dropping policies depends on the pricing strategy used. For example, while *LatestDeparture* is the best for *Prix Fixe 1* (as indicated previously), it performs no better than *Random* for the multi-class pricing strategies in Figure 4.

These results demonstrate the tradeoffs between call dropping policies and pricing strategies in stochastic capacity networks. In particular, the choice of an optimal pricing strategy depends on the call dropping policy implemented in the network, and vice versa.

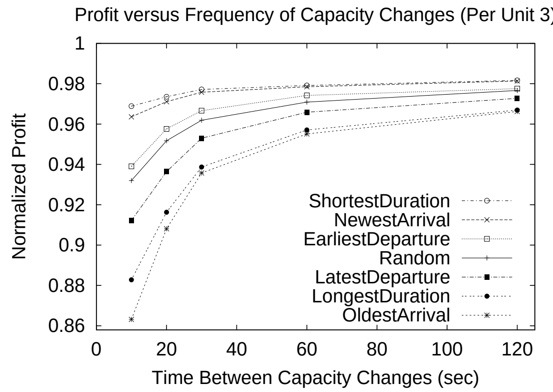
6 Conclusions

This paper studies possible pricing strategies for networks with stochastic capacity variation. Call-level simulation is used to study the performance tradeoffs between 4 pricing strategies and 7 different call dropping policies, with respect to their user-oriented and network-oriented performance. The simulations are conducted for a variety of call workload and network capacity assumptions.

The simulation results show that the choice of an optimal pricing strategy depends on the call dropping policy used in the network. For the simple pricing strategies studied in our experiments, the *NewestArrival* call dropping policy provides relatively robust performance, although it is rarely optimal. This simple-to-implement call dropping policy could be suitable as a default call dropping policy for stochastic capacity networks.



(a) Prix Fixe 3



(b) Per Unit 3

Figure 4. Effect of Frequency of Capacity Changes on Profit (100,000 calls; PE30; DN(40,5))

7 Acknowledgements

Financial support for this research was provided by iCORE (Informatics Circle of Research Excellence) in the Province of Alberta and by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The authors thank the anonymous reviewers for their constructive comments on our work.

References

[1] R. Akl and A. Parvez, "Impact of Interference Model on Capacity in CDMA Cellular Networks", *Proceedings of 8th World Multi-Conference on Systemics, Cybernetics, and Informatics (SCI)*, July 2004.

[2] W. Choi and J. Kim, "Forward Link Capacity of a DS/CDMA System with Mixed Multirate Sources", *IEEE Transactions on Vehicular Technology*, Vol. 50, No. 3, pp. 737-749, May 2001.

[3] S. Dharmaraja, K. Trivedi, and D. Logothetis, "Performance Modeling of Wireless Networks with Generally Distributed Handoff Interarrival Times", *Computer Communications*, Vol. 126, No. 15, September 2003.

[4] K. Gilhousen, I. Jacobs, R. Padovani, A. Viterbi, L. Weaver, and C. Wheatley, "On the Capacity of a Cellular CDMA System", *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, pp. 303-312, May 1991.

[5] A. Greenberg, R. Srikant, and W. Whitt, "Resource Sharing for Bookahead and Instantaneous-Request Calls", *IEEE/ACM Transactions on Networking*, Vol. 7, No. 1, pp. 10-22, February 1999.

[6] P. Gupta and P. Kumar, "The Capacity of Wireless Networks", *IEEE Transactions on Information Theory*, Vol. 46, No. 2, pp. 388-404, March 2000.

[7] B. Haverkort, R. Marie, R. Gerardo, and K. Trivedi, *Performability Modelling: Techniques and Tools*, Wiley, New York, 2001.

[8] W. Jeon and D. Jeong, "Call Admission Control for CDMA Mobile Communications Systems Supporting Multimedia Services", *IEEE Transactions on Wireless Communications*, Vol. 1, No. 4, pp. 649-659, October 2002.

[9] I. Koo, S. Bahng, and K. Kim, "Resource Reservation in Call Admission Control Schemes for CDMA Systems with Non-Uniform Traffic Distribution among Cells", *Proceedings of IEEE Vehicular Technology Conference (VTC)*, Vol. 1, pp. 22-25, April 2003.

[10] T. Shu and Z. Niu, "A Novel Call Admission Control Scheme for Multimedia CDMA Networks with Power Multiplexing and Imperfect Power Control", *Proceedings of Vehicular Technology Conference*, April 2003.

[11] J. Siwko and I. Rubin, "Call Admission Control for Capacity-Varying Networks", *Telecommunications Systems*, Vol. 16, No. 1-2, pp. 15-40, January 2001.

[12] J. Siwko and I. Rubin, "Connection Admission Control for Capacity-Varying Networks with Stochastic Capacity Change Times", *IEEE/ACM Transactions on Networking*, Vol. 9, No. 3, pp. 351-360, June 2001.

[13] H. Sun and C. Williamson, "Simulation Evaluation of Call Dropping Policies for Stochastic Capacity Networks", *Proceedings of the SCS Symposium on the Performance Evaluation of Computer and Telecommunication Systems* (SPECTS 2005), Philadelphia, PA, pp. 327-336, July 2005.