Simulation Evaluation of Call Dropping Policies for Stochastic Capacity Networks

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

Abstract

This paper studies the performance of call dropping policies for networks with stochastic capacity variation. Call-level simulation is used to compare the call blocking performance of 9 call dropping policies under a wide range of call workload and network capacity assumptions. Contrary to prior claims in the literature, our results show significant performance differences among dropping policies: the choice of an appropriate call dropping policy can significantly reduce call blocking, improving overall call-level performance in a stochastic capacity network. The effect is even more pronounced as the frequency and variance of capacity changes increase relative to the call workload characteristics. We propose an "equivalent capacity" model to capture the dynamics of these interactions.

Keywords: Stochastic Capacity Networks, Simulation, Call Blocking, Call Dropping, Wireless CDMA

1 Introduction

In many network environments, the available network capacity varies unpredictably with time [2, 5, 14, 15, 23]. For example, in a reservation-based network with multiple priority levels, high priority calls such as video conferences or emergency services may take precedence over ordinary traffic. The network capacity available for low priority traffic thus varies with time based on high priority traffic demands. In wireless networks, capacity variation arises from the mobility of users (e.g., handoffs) and the time-varying characteristics of the wireless propagation environment. The patterns of wireless interference for the active connections may dynamically change the available capacity for these connections [6, 8]. This phenomenon applies to wireless LANs and CDMA systems.

The traditional approaches to the limited-capacity problem are advance reservations, admission control, and adaptive rate control. For example, Asynchronous Transfer Mode (ATM) networks use Connection Admission Control (CAC) to determine which calls are allowed into the network, based on their requested resource demands and the currently available network capacity. To provide guaranteed Quality of Service (QoS), high priority calls must reserve resources in advance [3, 4, 19], or preempt resources at the time they are needed. In wireless CDMA systems, dynamic power control and rate adaptation are used to reduce the aggregate data rate when capacity problems occur [13, 21]. This approach maintains all active calls, but with degraded service quality.

In this paper, we study a different approach to the stochastic capacity problem, namely call dropping. This approach removes (drops) selected calls from the network when the traffic demands exceed the available capacity. In most networks, call dropping is viewed as a last resort. That is, call blocking at the time of call arrival is deemed preferable to call dropping in the middle of a call [16, 23]. Even if call dropping is permitted, the belief is that the call dropping policy makes little or no difference [7].

Our paper makes three main contributions. First, we show via simulation that call dropping policies can have an important impact on call-level performance in a stochastic capacity network. These differences manifest themselves at medium-to-high network load, with call blocking rates of 1-3%. In this regime, a well-chosen call dropping policy can reduce the overall call blocking rate by 16%. Second, we show that the choice of a call dropping policy is especially important in networks with high-frequency and high-variance capacity changes, as is typical in wireless CDMA systems with mobile users. In general, policies that reclaim more network resources can better tolerate high-frequency capacity changes. Finally, we propose an *equivalent capacity* model to reflect the interactions between the call workload and the stochastic capacity variation.

The rest of the paper is organized as follows. Section 2 briefly discusses prior related work. Section 3 defines our stochastic capacity models and call dropping policies. Section 4 describes the setup for our call-level simulation study, while Section 5 presents our simulation results. Section 6 discusses equivalent capacity modeling. Section 7 concludes the paper.

2 Related Work

There are many papers in the literature discussing CAC in variable-capacity networks [7, 9, 10, 11, 21, 22]. However, much of this work has restrictive assumptions about the call workload or the capacity variation process. Some CAC schemes assume a Poisson request arrival process, which may not be representative of realistic network traffic. Some approaches unrealistically assume advance knowledge about the capacity changes, or with only the timing of the changes considered as a stochastic process [23]. Other work [17, 23] models random capacity changes, but does not consider time correlations in the capacity variation.

These assumptions are worrisome because they may not be representative of capacity-varying networks. For example, in wireless data networks, the connection arrival process may not be Poisson [5, 12, 18], particularly when handoffs occur. Furthermore, capacity changes are not always known in advance, and there are many possible correlations induced by the active traffic patterns in the network [1, 13, 20].

Some papers in the literature state (or assume) that call dropping policies have little or no impact on the network performance [7]. Our work shows that this assumption is not true. To the best of our knowledge, our work is the first to explore this issue in detail.

One paper that does mention dropping policies is the paper by Siwko and Rubin [23]. Their paper focuses on CAC strategies in stochastic capacity networks, assuming IFR (increasing failure rate) holding time distributions. They consider a hybrid dropping policy combining Last-Come-First-Drop and Random, and then evaluate CAC strategies for network loads ranging up to 500%. By contrast, our work considers 9 different dropping policies at moderate-to-high network load, under the assumption of Greedy CAC.

Our work explores call-level performance in stochastic capacity networks, for a very broad range of call workload and network capacity assumptions. We argue that disrupting some calls is unavoidable in such a network when capacity decreases occur. An appropriate dropping policy is important to maintain overall system performance, particularly when the traffic demands and the capacity variation have complex statistical behaviours.

3 System Model

3.1 Call Workload

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: we consider Poisson as well as burstier arrival processes. Each call has a specified holding time, drawn from a specified distribution (e.g., Exponential, Pareto). Each call requires one unit of network capacity for the duration of the call.

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

3.2 Network Capacity 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.

An example of our stochastic network capacity model is shown in Figure 1. The horizontal axis represents time, while the solid line portrays the available network capacity at each instant in time. We model capacity changes as events that occur at specified points in 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 four characteristics of the stochastic capacity process. One characteristic is the frequency of capacity changes in the network. A second characteristic is the distribution used for the elapsed time between network capacity changes. We consider Deterministic, Exponential, and Self-Similar models for this timing structure. A third characteristic in the model is the distribution used for the network capacity itself. We focus primarily on the Normal distribution, for which it is easy to control both the mean and the variance of the network capacity. The mean of this distribution matches the long-term average of C calls, while the variance affects the magnitude of capacity fluctuations that can occur. A final characteristic is the correlation structure in the capacity time series

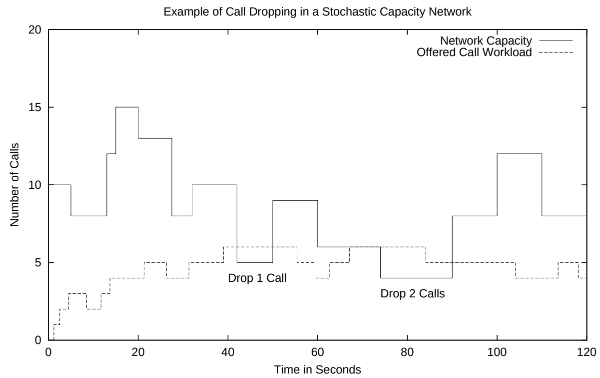


Figure 1. Stochastic Capacity Example

process. We consider independent and identically distributed (iid) samples as well as self-similar processes.

3.3 Call Dropping Policies

Call dropping occurs in a stochastic capacity network when the aggregate traffic demand 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. The network must expunge one or more *victim* calls in order to meet the new capacity constraints.

Two call dropping episodes are shown in Figure 1. The dashed line shows the offered call workload. It is a stochastic process, determined by call arrival and call departure events. The dashed line represents the ideal call occupancy in an infinite-capacity network, for this example workload. At time 42, a capacity decrease from $C = 10$ to $C = 5$ occurs, shortly after a sixth active call was admitted to the network. One victim call has to be dropped (removed) from the network at this point. At time 74, another call dropping episode occurs. This episode may require 2 additional calls to be dropped, perhaps depending on how the first call dropping episode was handled.

Choosing which call(s) to drop in a dropping episode is determined by a call dropping policy. In this paper, we consider 9 call dropping policies, in five categories:

- **Randomized Policies**: The first policy considered is the *Random* policy. It chooses uniformly at random amongst the active calls in the network whenever a victim call must be dropped. This simple policy provides a baseline for comparison of other policies. Policies that perform worse than *Random* are undesirable.

- **Arrival-based Policies**: The arrival-based policies use call arrival time information to determine the 'age' of each existing call. Two such policies are considered. The *NewestArrival* policy removes the youngest call from the network, enforcing a Last-In-First-Out (LIFO) policy when call dropping occurs. The *OldestArrival* policy removes the oldest call from the network, enforcing a First-In-First-Out (FIFO) policy. Both policies rely only on the past history of call arrivals, and are easy to implement in practice.

- **Departure-based Policies**: The departure-based policies rely on call departure time information. In a general network, these policies require omniscient future knowledge, and are not practically implementable. In a reservation network, the call departure time is determined from the call arrival time and the reservation duration. Two different policies are considered. The *EarliestDeparture* policy removes from the network the active call that is scheduled to complete next. The *LatestDeparture* policy drops the active call whose departure time is furthest in the future.

- **Duration-based Policies**: These policies use call duration information only. The *ShortestDuration* policy removes from the network the call with the shortest holding time specified at the time of its arrival, regardless of its current state. The *LongestDuration* policy drops the active call with the longest original holding time, regardless of its current state. In a general network, these policies require future knowledge, or good heuristics for estimating call duration.

- **Completion-based Policies**: The completion-based policies use a combination of call arrival time and call duration information. Calls are ranked based on their relative closeness to completion, expressed as a percentage. The *LeastCompleted* policy removes from the network the call that has completed the smallest proportion of its originally-intended service. The *MostCompleted* policy drops the call that has completed the largest proportion of its planned service.

4 Experimental Methodology

In the rest of the paper, we use call-level simulation to illustrate the performance differences among call dropping policies. The simulation approach allows us to study a wide range of call workload and network capacity characteristics, generalizing our observations to more realistic network assumptions. This section describes the experimental setup for our simulation study, while Section 5 presents the simulation results.

4.1 Simulation Model

Our work is carried out using call-level simulation. We model admission control at an access node in a network with time-varying capacity.

The two inputs provided to the simulation are a call workload file and a network capacity file. The call workload file is a time-ordered sequence of call arrival events. Each call specifies its source node, destination node, arrival time, and duration. Each call requires one unit of network capacity. Workload files are generated using the call workload models indicated in Table 1. We use workload files with 100,000 calls. We consider this trace length adequate to highlight performance differences among the call dropping policies evaluated.

Table 1. Call-Level Workload Parameters

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

The network capacity file is a time-ordered sequence of capacity change events. Capacity files are generated using the models and parameters indicated in Table 2. We use capacity files with 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.

The different call dropping policies are modeled within the simulator. We provide each policy with the same workload and capacity files, so that they each handle the same traffic demands under the same network conditions. Differences observed in the call-level performance reflect differences in the call dropping policies used. Only the *Random* policy has non-deterministic (randomized) behaviour. For this policy, we use 3 runs with different random number seeds.

4.2 Experimental Design

Table 3 shows the factors and levels used in our simulation experiments. We explore the impact of different call dropping policies on the call-level performance, for a broad range of network capacity variation.

The primary performance metrics are the call blocking probability and the call dropping probability. These metrics characterize the user-perceived QoS.

5 Simulation Results

5.1 Overview

Figure 2 provides an overview of the simulation results, by plotting call-level performance as a function of offered load. These results are from a single representative simulation run with 100,000 calls. Figure 2(a) presents the call blocking results, showing the proportion of offered calls rejected from the network at the time of their arrival. Figure 2(b) presents the call dropping results, showing the proportion of the accepted calls that are subsequently dropped from the network prior to their completion. In both plots, the horizontal axis shows the offered load in Erlangs. The call arrival process is Poisson, and the call holding times are exponentially-distributed with a mean of 30 seconds. 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 2. We chose these parameter values as a model for a typical commercial wireless CDMA system.

Figure 2 shows that the call blocking rate and the call dropping rate both increase with offered load, as expected. Call blocking and call dropping are negligible at or below a load of 20 Erlangs. It is rare for the network to approach saturation at this load level, and rare for a capacity change to trigger call dropping. When the offered load is 30 Erlangs (75% average load), call blocking becomes noticeable (about 2%). A small proportion of call dropping also occurs (about 0.4%). Beyond a load of 40 Erlangs, the network is overloaded, and the call blocking rate exceeds 10%. The call dropping rate increases as well.

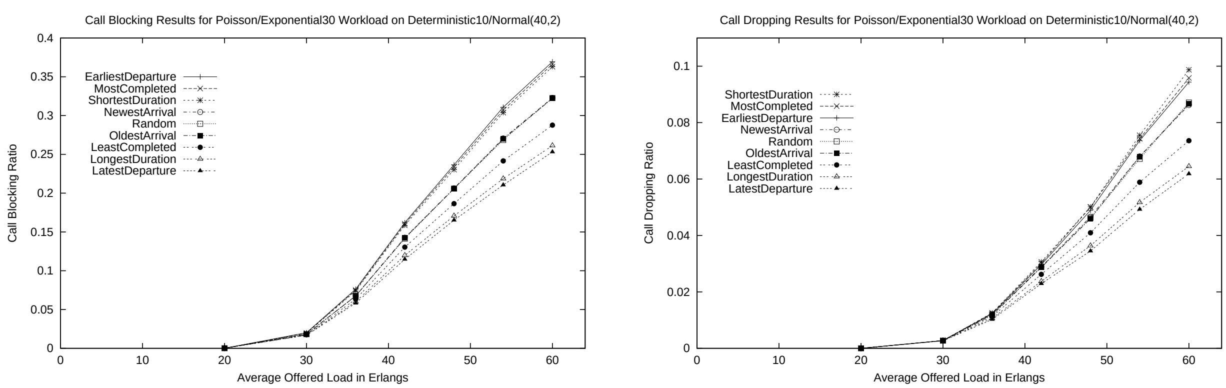
Three main observations are evident from Figure 2. First, the simulation results are structurally similar for both call blocking and call dropping, though the call dropping rate is typically much lower than the call blocking rate (note the different vertical scales on the graphs). This property holds throughout our simulation results. Second, performance differences between the 9 dropping policies become apparent, especially at higher loads. The best policies are *LatestDeparture*, *LongestDuration*, and *LeastCompleted*, which tend to reclaim a relatively large amount of (previously committed) network capacity when they disrupt a call. The worst policies are *EarliestDeparture*, *ShortestDuration*, and *MostCompleted*, which do not. The in-between policies are *Random*, *NewestArrival*, and *OldestArrival*. In fact, the simulation results show that arrival-based policies (*NewestArrival* and *OldestArrival*) are no more effective than *Random* call dropping. Third,

Table 2. Network Capacity Parameter Settings in Call-Level Simulations

Parameter		Levels
Mean Time between Capacity Changes (sec)		10, 15, 30, 60, 120
Stochastic Capacity	Capacity Change Time	Deterministic, Exponential, Self-Similar
	Capacity Change Value	Normal
Capacity Value (calls)	Mean	40
	Standard Deviation	2, 5

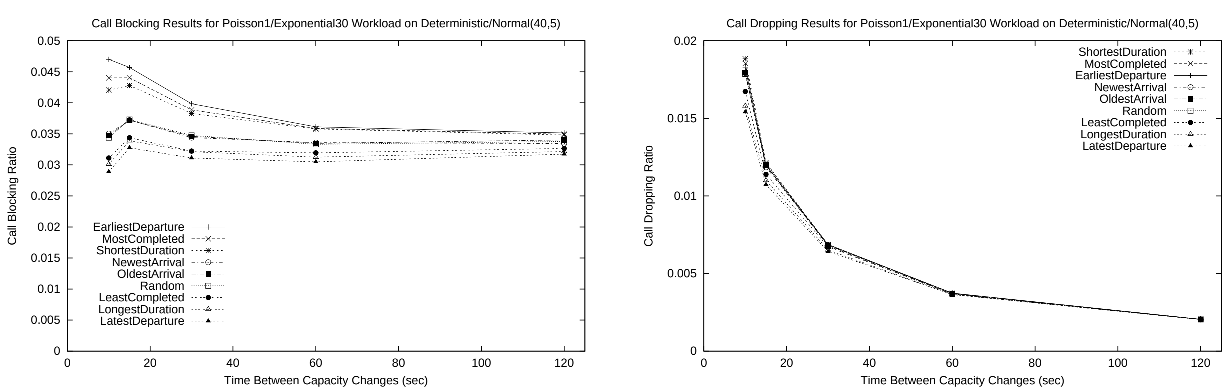
Table 3. Factors and Levels in Call-Level Simulations

Factor		Levels
Call Dropping Policy		Random, NewestArrival, OldestArrival, EarliestDeparture, LatestDeparture, ShortestDuration, LongestDuration, LeastCompleted, MostCompleted
Stochastic Capacity	Capacity Change Time	Deterministic, Exponential, Self-Similar
	Capacity Change Value	Normal



(a) Call Blocking Results (b) Call Dropping Results

Figure 2. Simulation Results for Call-Level Performance as a Function of Offered Load



(a) Call Blocking Results (b) Call Dropping Results

Figure 3. Effect of Frequency of Capacity Changes on Call-Level Performance

the relative ordering of policies in Figure 2(a) is quite consistent with that in Figure 2(b). In both graphs, the key shows the relative ordering of policies at a load of 60 Erlangs. Policies with a low call dropping rate tend to have a low call blocking rate.

These observations motivate our detailed study of call dropping policies in the following subsections. Unless stated otherwise, all remaining experiments use an offered load of 30 Erlangs, so that the average call blocking rate is about 1-3%. We consider this regime more relevant than the heavy overload conditions studied by Siwko and Rubin [23].

5.2 Frequency of Capacity Changes

Next, we vary the frequency of capacity changes in the network, to study the effect on call-level performance. Figure 3 shows these simulation results, with call blocking in Figure 3(a) and call dropping in Figure 3(b). In both plots, the horizontal axis represents the time between random capacity changes in the network. For these results, the mean network capacity is drawn from a Normal distribution, with a mean of 40 calls and a standard deviation of 5. The capacity changes exactly every T seconds, where T is indicated along the axis. The left end of the scale represents high frequency changes (every 10 seconds), while the right end represents low frequency changes (every 120 seconds). The mean call holding time is 30 seconds.

Figure 3 illustrates six important observations. First, the average call blocking rate (3.5%) is slightly higher here than in Figure 2(a) for a load of 30 Erlangs. This result occurs because of the higher standard deviation (5 versus 2) in the stochastic network capacity model. Second, the number of calls dropped depends on the frequency of capacity changes. As the time between capacity changes increases, the call blocking rates for all policies asymptotically converge toward the same value, and the call dropping rate asymptotically approaches 0. This result is as expected, since low-frequency changes approximate a static network, for which the Erlang B blocking formula can be directly applied [24]. If capacity changes are infrequent, few calls need to be disrupted. Third, the performance differences between dropping policies are more pronounced when there is a high frequency of capacity changes in the network. This result makes sense since high-frequency changes imply more call dropping episodes, and thus greater opportunity for distinctions among policies. Fourth, the differences among policies manifest themselves more clearly in the call blocking performance than in the call dropping performance. The negligible differences in call dropping performance indicate that all policies disrupt about the same number of calls. The important observation is that carefully choosing *which* calls are disrupted can significantly benefit the call blocking performance. For example, for high-frequency capacity changes, the call blocking rate is 2.89% for the *LatestDeparture* policy (16% lower than for the *Random* policy) and 4.70% for the *EarliestDeparture* policy (37% higher than for the *Random* policy). Fifth, the relative ordering of policies here is consistent with that in Figure 2. *LatestDeparture*, *LongestDuration*, and *LeastCompleted* provide the best performance, while *EarliestDeparture*, *ShortestDuration*, and *MostCompleted* provide the worst. Finally, the relationship between call blocking rate and frequency of capacity changes in Figure 3(a) is not strictly monotonic. For some policies, the call blocking rate decreases as capacity changes become less frequent, while for other policies, including *Random*, the behaviour is non-monotonic. This phenomenon is discussed further later in the paper.

5.3 Variability of Capacity Changes

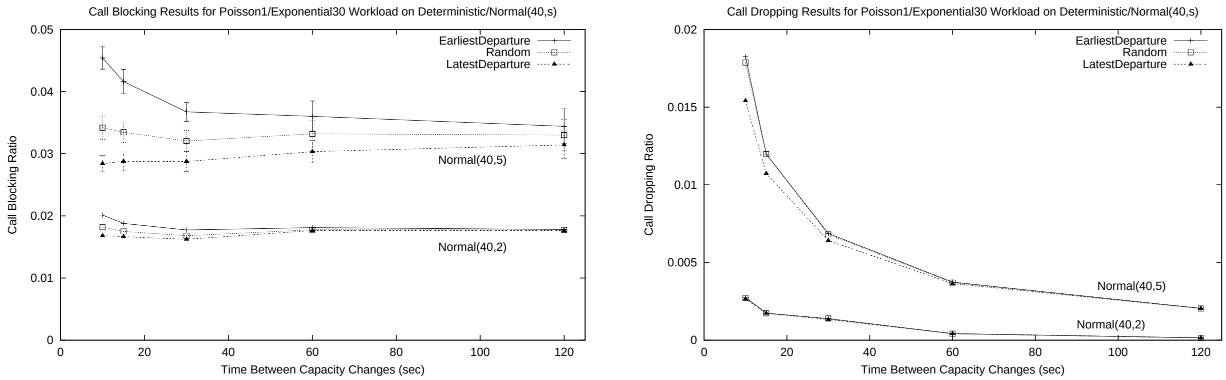
Figure 4 shows the impact of network capacity variability on call-level performance. These graphs show call blocking (Figure 4(a)) and call dropping (Figure 4(b)) results for three selected dropping policies, namely *EarliestDeparture* (worst), *Random*, and *LatestDeparture* (best). Results are shown for two different stochastic capacity models. Each has capacity changes every T seconds ($10 \leq T \leq 120$), with the capacity value drawn from a Normal distribution with a mean of 40 calls. However, one has a standard deviation of 2, and the other a standard deviation of 5. The vertical error bars in Figure 4(a) show the minimum and maximum call blocking rates observed from 10 simulation runs. Error bars are omitted on other graphs to improve readability.

Figure 4 shows that the higher-variability capacity model has a higher call blocking rate and a higher call dropping rate. These results are as expected, because more capacity variation occurs. The separation between dropping policies is more pronounced with higher capacity variability.

These results show that for networks with high-frequency or high-variability capacity changes, the call dropping policy can have a large impact on call blocking performance.

5.4 Timing of Capacity Changes

Figure 5 shows the impact of the capacity change process on the call-level performance. We consider



(a) Call Blocking Results (b) Call Dropping Results

Figure 4. Effect of Capacity Variability on Call-Level Performance

three different models for the timing between capacity change events: Deterministic, Exponential, and Self-Similar. The Deterministic model has a capacity change event every T seconds. The Exponential model has capacity change events at random times, following a Poisson process. The time between capacity change events is exponentially distributed with a mean of T seconds. The Self-Similar model assumes that capacity change events occur in a bursty fashion, similar to a self-similar (fractal) process. The mean time between capacity change events is T seconds.

Figure 5 shows that the timing structure of the capacity change process has a small impact on the call blocking performance (Figure 5(a)), but a larger impact on the call dropping performance (Figure 5(b)). These results are for a Normal(40,2) distribution of the network capacity. Figure 5(a) shows a complex relationship between the call blocking rate and the frequency of capacity changes. These results are for two dropping policies (*LatestDeparture* and *Random*), and the three different capacity change models.

5.5 Correlation of Capacity Changes

Figure 6 studies the effect of correlations in the capacity values. These results consider capacity changes every 30 seconds for two different capacity value models: Self-Similar and Random. In the Self-Similar model, the capacity values constitute a self-similar process, with short-range and long-range correlations. In the Random model, the same capacity trace is shuffled into a random order to remove short-range and long-range correlations. In both models, the mean and standard deviation of the capacity are the same.

Figure 6 shows that correlations in the capacity change process are beneficial. The upper three lines in each plot are for the Random (uncorrelated) model, which can have large fluctuations in network capacity at any time scale. The lower three lines are for the Self-Similar model, which has both short-term and long-term correlation present in the capacity value time series. Correlated capacity values produce more gradual changes in capacity, with less severe call blocking and call dropping rates. In other words, the presence of correlation in the capacity value time series changes the "equivalent capacity" of the network.

6 Equivalent Capacity Modeling

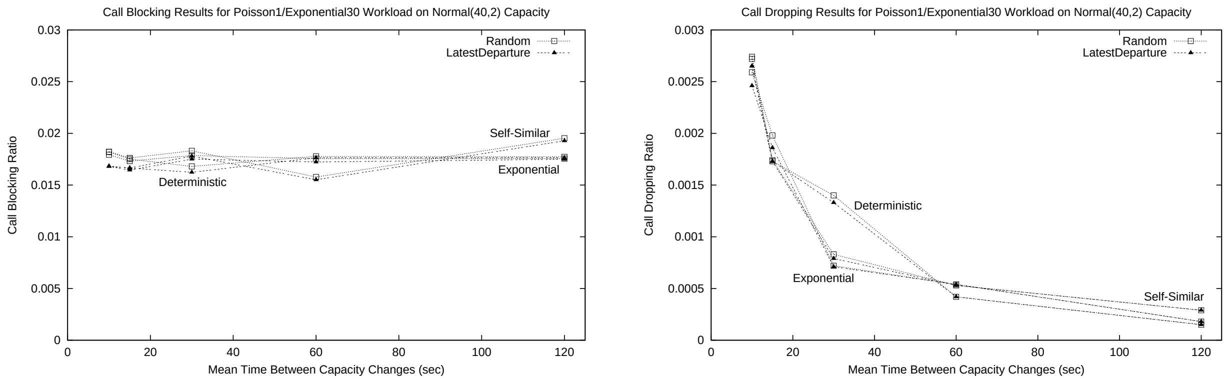
This section discusses the concept of *equivalent capacity*, and the interplay between call workloads and stochastic network capacity variation.

6.1 Motivation

The classic Erlang B formula expresses the relationship between offered load (in Erlangs), link capacity, and call blocking rate [24]. Given any two of these values, the third can be calculated directly. An illustration of this relationship appears in Figure 7(a).

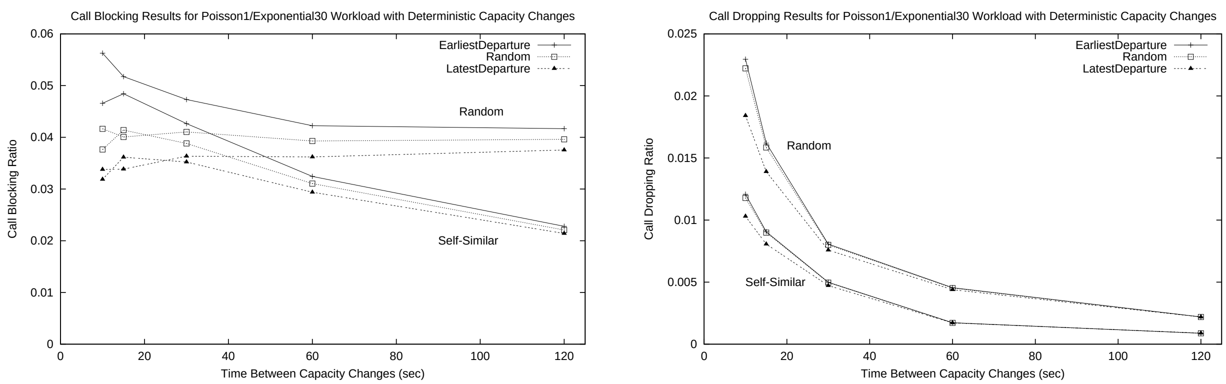
There are two main assumptions in the derivation of the Erlang B formula. The first assumption is that the link capacity is fixed. This finite capacity bounds the one-dimensional Markov chain used in the derivation, so that the loss rate can be determined. The second assumption is that the call arrival process is Poisson, to facilitate Markov chain analysis. Some authors further assume exponentially-distributed call holding times, but this is not strictly required. Only the mean holding time matters.

In stochastic capacity networks, the first assumption is violated, so the Erlang B formula no longer applies. Therefore, other solutions are required. The most obvious approach is a two-dimensional Markov chain, with



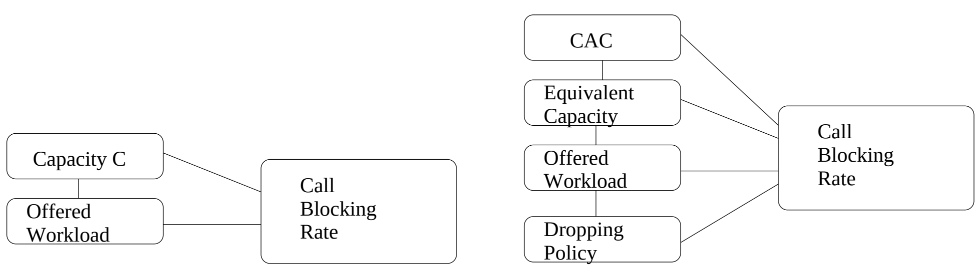
(a) Call Blocking Results (b) Call Dropping Results

Figure 5. Effect of Burstiness of Capacity Change Process on Call-Level Performance



(a) Call Blocking Results (b) Call Dropping Results

Figure 6. Effect of Capacity Change Correlations on Call-Level Performance



(a) Erlang B Modeling Approach (b) Equivalent Capacity Modeling Approach

Figure 7. Illustration of Modeling Framework

the second dimension modeling the network capacity. However, this requires an assumption that the capacity variation process is Markovian, which restricts the scope of the analysis.

Our ongoing work is exploring approaches for performance modeling and analysis of stochastic capacity networks. Analogous to the *Equivalent Random Traffic* methods adapting non-Markovian traffic to the Erlang B formula, we envision an *Equivalent Capacity* model to account for stochastic capacity variation. Our current simulation work makes preliminary contributions toward this goal.

6.2 Modeling Framework

Figure 7(b) shows the conceptual framework for our equivalent capacity modeling. We believe that three different inputs are required compared to Figure 7(a). The first of these new inputs is the stochastic capacity variation process. The second is the call dropping policy. The third is CAC. We comment briefly on each of these below.

As demonstrated by our simulation results, stochastic capacity variation can have a dramatic impact on call-level performance. The important factors are the frequency and variance of the capacity changes, especially their behaviour relative to the call workload characteristics. High frequency capacity changes, relative to the call holding times, lead to a high call dropping rate. High variability in the stochastic capacity process, relative to the variability of the call workload, strongly influences the call-level performance.

The call dropping policy and CAC also have an important role to play. Across a wide range of offered load, there are noticeable differences in the call-level performance with different call dropping policies. These differences manifest themselves at medium-to-high load, with call blocking rates of 1-3%.

The main reason that call dropping policies matter is that they affect resource reclamation, and thus the equivalent network capacity. They also affect the relative time scales of the capacity variation process. Dropping policies thus need to be part of the equivalent capacity model.

We have made some progress on the mathematical modeling of equivalent capacity, but the results are too preliminary to include here. Instead, we close our discussion with an illustrative example of the influences of call dropping policies and equivalent capacity on call blocking performance.

Table 4. Call-Blocking Simulation Results for Workload PE30 on Capacity DN(40,2)

Dropping Policy	Mean Time Between Capacity Changes				
	10 s	15 s	30 s	60 s	120 s
Random	1.814	1.740	**1.685**	1.777	1.773
NewestArr	1.819	1.746	**1.679**	1.774	1.774
OldestArr	1.829	1.752	**1.682**	1.774	1.773
EarliestDep	2.013	1.879	**1.773**	1.821	1.781
LatestDep	1.682	1.665	**1.623**	1.764	1.764
ShortestDur	1.939	1.826	**1.744**	1.805	1.776
LongestDur	1.707	1.685	**1.635**	1.767	1.769
LeastComp	1.758	1.701	**1.655**	1.767	1.771
MostComp	1.974	1.851	**1.762**	1.805	1.780

6.3 Numerical Example

There is a subtle but interesting phenomenon in our simulation results, wherein the call blocking rate exhibits non-monotonic behaviour with respect to the frequency of capacity changes. Each call dropping policy exhibits its own behaviour, often with the lowest call blocking rate achieved at a different capacity change frequency than other policies.

We investigated this phenomenon further, to understand whether it is a statistical anomaly, or an inherent behaviour. We believe that this phenomenon reflects the interactions between the stochastic traffic workload and the stochastic network capacity. The non-monotonic behaviour is related to the relative time scales of the two stochastic processes.

Additional evidence of this phenomenon appears in Table 4 and Table 5. These tables show the mean call blocking rate from the simulations for two different capacity change timing models: Deterministic in Table 4, and Exponential in Table 5. Both assume a Normal(40,2) distribution for the capacity values. Results are shown for all 9 call dropping policies, and 5 different time scales for capacity changes.

The lowest call blocking rate (highlighted in bold) for each dropping policy is observed in the middle column of Table 4, where the mean time between capacity changes (30 sec) matches the mean call holding time. In separate experiments varying the mean call holding time (not shown here), the lowest call blocking rate tends to shift columns accordingly. That is, call blocking is lowest when there is an average of 1 capacity change event per call duration.

Surprisingly, *none* of the policies achieve their lowest blocking rate in the middle column of Table 5. The only difference here is the random timing of the capacity

Table 5. Call-Blocking Simulation Results for Workload PE30 on Capacity EN(40,2)

Dropping Policy	Mean Time Between Capacity Changes				
	10 s	15 s	30 s	60 s	120 s
Random	1.794	**1.729**	1.789	1.758	1.756
NewestArr	1.808	**1.735**	1.792	1.757	1.758
OldestArr	1.800	**1.724**	1.790	1.755	1.754
EarliestDep	1.974	1.871	1.834	1.797	**1.773**
LatestDep	1.680	**1.645**	1.751	1.723	1.749
ShortestDur	1.922	1.825	1.821	1.789	**1.767**
LongestDur	1.711	**1.660**	1.761	1.729	1.750
LeastComp	1.740	**1.687**	1.777	1.736	1.750
MostComp	1.942	1.849	1.826	1.791	**1.772**

changes (Exponential instead of Deterministic). For some policies, the lowest call blocking rate moves to the left (i.e., higher frequency changes, at shorter time scales), while for other policies, the lowest call blocking rate moves to the right (i.e., lower frequency changes, at longer time scales).

In general, policies that reclaim more network resources (i.e., *LatestDeparture*, *LongestDuration*, and *LeastCompleted*) can better tolerate high-frequency capacity changes. This phenomenon contributes to the non-monotonicity observed. It also illustrates the non-trivial interplay between dropping policies, call workload, and stochastic capacity variation.

7 Conclusions

This paper studies call dropping policies for networks with stochastic capacity variation. Call-level simulation is used to study 9 different call dropping policies with respect to their call blocking and call dropping rates. The simulations are conducted for a broad set of call workload and network capacity assumptions.

The results show significant differences among dropping policies, particularly for high-frequency and high-variability capacity change models. The choice of an appropriate call dropping policy can reduce the call blocking rate significantly, improving overall call-level performance in a stochastic capacity network.

Ongoing work is refining our mathematical analysis of stochastic capacity networks using a semi-Markov model. Future work will study CAC schemes for stochastic capacity networks, as well as heuristics for classifying flow durations. Stochastic capacity models derived from wireless user mobility are also under development.

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] F. Ashtiani, J. Salehi, and M. Aref, "A Flexible Dynamic Traffic Model for Reverse Link CDMA Cellular Networks", *IEEE Transactions on Wireless Communications*, Vol. 3, No. 1, pp. 159-164, January 2004.

[3] N. Bortolini, "Handoff and Optimal Channel Assignment in Wireless Networks", *Mobile Networks and Applications*, Vol. 6, No. 6, pp. 511-524, November 2000.

[4] H. Chen, S. Kumar, and C. Kuo, "Dynamic Call Admission Control Scheme for QoS Priority Handoff in Multimedia Cellular Systems", *IEEE Wireless Communications and Networking Conference*, Orlando, FL, March 2002.

[5] 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.

[6] 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.

[7] 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.

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

[9] A. Hernandez, A. Valdovinos, and F. Casadevall, "Capacity Analysis and Call Admission Techniques for CDMA Packet Transmission Systems", *Proceedings of 4th IEEE International Workshop on Mobile and Wireless Communications Networks*, Stockholm, Sweden, September 2002.

[10] A. Iera, A. Molinaro, E. Natalizio, and S. Marano, "Call Management Based on the Mobile Terminal-Peak Velocity: Virtues and Limitations in a Two-

Tier Cellular System", *IEEE Transactions on Vehicular Technology*, Vol. 52, No. 4, pp. 794-813, July 2003.

[11] 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.

[12] S. Kasahara, "Internet Traffic Modeling: Markovian Approach to Self-Similar Traffic and Prediction of Loss Probability for Finite Queues", *IEICE Transactions on Communications*, Vol. E84-B, No. 8, August 2001.

[13] D. Kim and D. Sung, "Capacity Estimation for a Multicode CDMA System with SIR-based Power Control", *IEEE Transactions on Vehicular Technology*, Vol. 50, No. 3, pp. 701-710, May 2001.

[14] 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.

[15] B. Li, L. Li, B. Li, K. Sivalingam, and X. Cao, "Call Admission Control for Voice/Data Integrated Cellular Networks: Performance Analysis and Comparative Study", *IEEE Journal on Selected Areas in Communications*, Special Issue on All-IP Wireless Networks, January 2004.

[16] W. Li, Y. Fang, and R. Henry, "Actual Call Connection Time Characterization for Wireless Mobile Networks Under a General Channel Allocation Scheme", *IEEE Transactions on Wireless Communications*, Vol. 1, No. 4, pp. 682-691, October 2002.

[17] Z. Liu and M. El Zarki, "SIR-Based Call Admission Control for DS-CDMA Cellular Systems", *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 4, pp. 638-644, May 1994.

[18] L. Muscariello, M. Mellia, M. Meo, R. Cigno, and M. Ajmone Marsan, "An MMPP-Based Hierarchical Model of Internet Traffic", *Proceedings of IEEE International Conference on Communications (ICC)*, Paris, France, June 2004.

[19] C. Redding and C. Taylor, "Priority Access Service in Cellular and PC Networks", *Proceedings of MILCOM*, October 2001.

[20] D. Shen and C. Ji, "Capacity Trade-offs for Heterogeneous Traffic over Large scale Fading Channels in CDMA Networks", *Proceedings of IEEE International Conference on Communications (ICC)*, Vol. 6, pp. 1718-1722, 2001.

[21] 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.

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

[23] 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.

[24] R. Wilkinson, "Theory for Toll Traffic Engineering in the USA", *Bell Systems Technical Journal*, Vol. 35, pp. 421-514, 1956.