Queueing model for soft-blocking CDMA systems

Lachlan L. H. Andrew, Donald J. B. Payne and Stephen V. Hanly
Department of Electrical and Electronic Engineering
University of Melbourne, Parkville, Vic 3052, AUSTRALIA
Ph +61 3 9344 9208 Fax +61 3 9344 9188
{L.Andrew,djbp,S.Hanly}@ee.mu.oz.au

Abstract— **A new model is proposed for the analysis of CDMA systems. This model is a birth and death process whose birth process considers the new call arrival rate, the blocking rate, the effect of soft handoff and the effect of the band allocation strategy in multiband (multicarrier) CDMA systems. This model accurately predicts the distribution of the number of calls connected to a base station.**

I. Introduction

Code division multiple access (CDMA) systems are characterised by "soft blocking". New calls are blocked when the total same cell plus other cell interference is too high. Thus they can neither be considered to be finite queues, which accept arrivals until a fixed threshold is reached, or infinite queues, which never block arrivals. The traditional M/M/∞ queue model [1] does not reflect this soft blocking behaviour of a CDMA system. This paper develops an alternative model, in which each cell blocks newly arriving calls with a state-dependent probability. In its simplest form this model is similar to that used (but not directly evaluated) in [2].

This model treats each cell of a multi-cell system independently. The state of the cell is the number of users currently in that cell. Other cells simply contribute random interference. This random interference causes blocking with some probability, B_i, which is assumed to depend only on the state i of the current cell.

The state of the cell can be modelled as a birth and death process [3]. Deaths (call departures) have negative exponential inter-event times with rate proportional to the current number of calls. Births also have a negative exponential inter-event time, with rate equal to the arrival rate thinned by the blocking probability, $\lambda(1 - B_i)$.

In this paper, B_i are determined by simulation. The question arises "if B_i are from simulation, why model the system?" Modelling each cell as independent, and treating all other cells as an independent random process is an approximation. If the simulation results agree with the Markov model, this validates that approximation. In future, closed form expressions for the B_is will be determined, and a fully analytic solution will be available.

This work was funded in part by the Australian Research Council (ARC)

II. Hard handoff

When hard handoff is used, new calls are allocated to the nearest base station. This gives rise to the very simple birth and death process with the arrival rate to state i being $\lambda(1 - B_i)$, and the departure rate $i\mu$, shown in Figure 1. Call this Markov chain "Markov chain A". Figure 2 shows the probability distribution for the state of a cell according to an event driven simulation of the system and the state distribution of Markov chain A. For comparison, it also shows the distribution for an M/M/∞ model. Clearly the proposed model gives a much better fit than the traditional model. This simulation was for a 4×4 hexagonal grid with a spreading factor of 128, and calls accepted if the signal to interference ratio (SIR) exceeds 6 dB. Log-normal shadowing was assumed, with standard deviation $\sigma = 8$ dB, but multipath fading was ignored. The load was 14 Erlangs. The overall blocking probability was 12.7%, both by simulation and the Markov A model. (Note that very high blocking probabilities are used throughout this paper to highlight the difference in the models.) When the blocking probability is low, it may be approximated by the probability that the number of calls exceeds a maximum permissible value in an M/M/∞ model. The permissible value typically used is $W/\alpha - 1 - (1 - B)f\lambda/\mu$, where W is the processing gain, α is the required SIR, B is the overall blocking probability, f is the ratio of same-cell to other-cell interference in a uniformly loaded sys-

0-7803-5435-4/99/$10.00 © 1999 IEEE

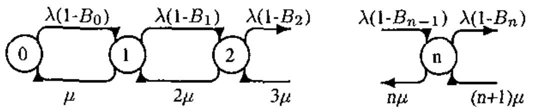

Fig. 1. Markov chain for hard handoff CDMA.

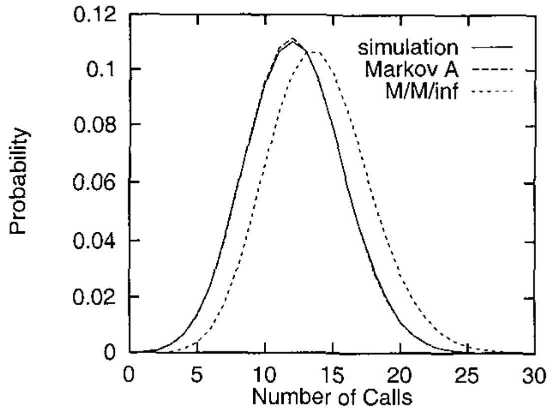

Fig. 2. State distribution for hard handoff: simulated, proposed Markov model and M/M/∞ model.

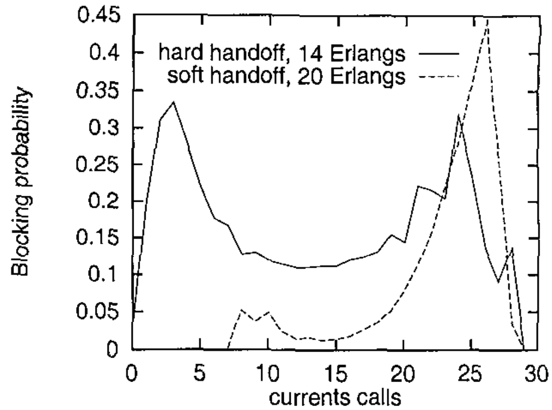

Fig. 3. Blocking probability in each state for the simulation used in Figure 2 and, for comparison, the blocking probability per state for a soft handoff system.

tem, and λ/μ is the load on the system. Since the threshold depends on the blocking probability, an iterative solution is required (although in many cases the approximation $B \approx 0$ suffices). For hard handoff, $f \approx 2.5$, and in this case the blocking probability is approximately 50%, the probability of exceeding approximately 14 calls. Clearly the approximation that the distribution is approximately that of an M/M/∞ queue is not valid in this case, and thus the result is substantially different from the true value and is of no use in network planning.

Figure 3 shows the blocking probability in each state for the above simulation. As expected, the blocking probability is greater for heavily loaded cells than for cells with a moderate load. A more unexpected result is that blocking is also higher for very *lightly* loaded cells. This is because of the correlation in blocking events. If the other-cell interference is high when a call arrives, it will probably also be high when the next call arrives. Thus when the other-cell interference is high for a prolonged period, the occupancy of the cell will drop as users leave but no new users are admitted. While it is surprising that blocking is high when the number of calls in a cell is low, the converse (that the number of calls in a cell is low when the blocking is high) is intuitively obvious. To verify this explanation, the figure also shows the blocking per state for soft handoff. (The load was increased to obtain per-state blocking of the same order of magnitude as for hard handoff, although the overall blocking was not matched.) In soft handoff, the other cell interference is greatly reduced since users are decoded by the base station at which they have the highest SIR. Hence the total interference will drop more markedly as the cell empties than is the case for hard handoff, and prolonged periods of high interference are less common. This phenomenon is not captured by the statistical models used in [2] and elsewhere.

III. Soft handoff

In soft handoff, a mobile connects simultaneously to several base stations, and is decoded by the base station with the highest SIR. This greatly improves system capacity. The state of a cell is then the number of users which are currently being decoded by that cell. Applying the blocking probabilities measured for soft handoff to Markov chain A gives a very poor fit to the soft handoff simulation. This can be seen in Figure 4, which shows the results for a load of 20 Erlangs. The overall blocking rate is 7.3%, which is the same order of magnitude as the simulated value of 5.6%, but that is largely due to the cancellation of the overestimate of blocking due to very heavily loaded cells by the underestimate of blocking due to moderately loaded cells.

Soft handoff produces a distribution with more cells having approximately the average number of calls than hard handoff. That is because calls will typically have a higher SIR in more lightly loaded cells, and thus will be more likely to connect to the corresponding base stations, leading to a higher effective arrival rate at base stations with fewer calls. This can be modelled by making the arrival rate depend on the current state of the cell. A suitable and effective heuristic for this arrival rate will now be outlined.

For this purpose it is reasonable to make the approximation that all of the neighbouring cells have an equal number of users n. Initially it will be assumed that $n = \lambda/\mu$. It is also reasonable to assume that the effective arrival rate will be proportional to the true arrival rate, and so the effective arrival rate in state i will be $k_{i,n}\lambda(1 - B_i)$. Similarly, $k_{i,n}$ should not change when the entire load on the system scales uniformly, so it will depend on i and n only through the ratio i/n. When $i/n \approx 1$, the system is uniformly loaded, and the arrival rate at all cells will simply be the true arrival rate, yielding $k_{i,n} \approx 1$. For $i/n \gg 1$, $k_{i,n}$ will be almost zero, since most users will connect to neighbouring base stations, rather than the base station of interest. Finally, for $i/n \ll 1$, $k_{i,n}$ will be around 2, since the base station will get all of the calls in its own cell, and those calls in the neighbouring cells for which the base station of interest is the second nearest. One functional form that meets these requirements is $2/(1 + (i/n)^a)$ for $a > 0$. Empirically it has been found that this provides a good fit with $a = 4$. Figure 5 shows this scaling factor against i/n, the occupancy of the current cell relative to the mean load. Thus the overall arrival rate is

$$\frac{2\lambda}{1 + (i\mu/\lambda)^4}(1 - B_i)$$

in state i. Call this modified Markov chain "Markov chain B". Once an estimate of the overall blocking probability, B, is known, the Markov chain can be improved by approximating the occupancy of neighbouring cells as $n \approx (\lambda/\mu)(1 - B)$, yielding "Markov chain B with two iterations". This may be repeated until the blocking probability converges. The results for soft handoff with a load of 20 Erlangs are shown in Figure 4. After one iteration, Markov chain B produced a much better approximation to the true state distribution than Markov chain A. However, it slightly overestimated the number of calls in the cell (that is, the mean of the distribution was shifted to the right), and it predicted an overall blocking rate of 7.4%, which is very similar to that of Markov chain A, but substantially different from the true value of 5.6% obtained by simulation. After a second iteration, the model produced a very good approximation to the simulated state distribution, and predicted a blocking probability of 5.4%. For comparison, the results with no blocking are included.

The fact that results from Markov chain B match the simulated curve well but those from Markov chain A do not shows that the good match is not "inevitable" given the per-state blocking probabilities, but rather that the choice of Markov chain is responsible for the good match.

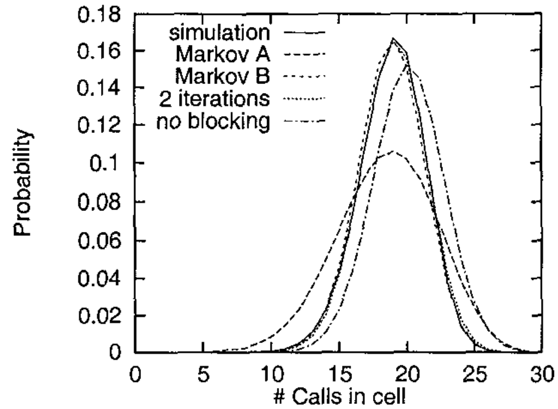

Fig. 4. State distribution for soft handoff: Simulated, original Markov chain, modified Markov chain.

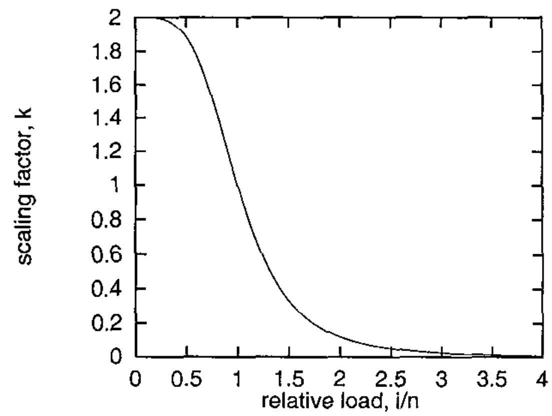

Fig. 5. Scaling factor applied to arrival rate, as a function of i/n.

Note that it was not necessary to alter the arrival rates in this way for soft handoff in [2]. That is because base station selection there was based on the measured path gain, rather than the SIR. Thus the base station to which a new call connected depended only on the propagation conditions and was independent of the load on the cell, causing the arrival rate (before blocking) to be indpendent of the state.

IV. Multi-band operation

The capacity of a CDMA system may be easily increased using multi-band (or multi-channel) CDMA, in which a second slice of spectrum is used, and calls are allocated to one or other band [4–6]. The performance of such systems is influenced by the way new calls are allocated to bands, and this difference in performance can be predicted by the proposed technique. The Markov chain will now be two dimen-

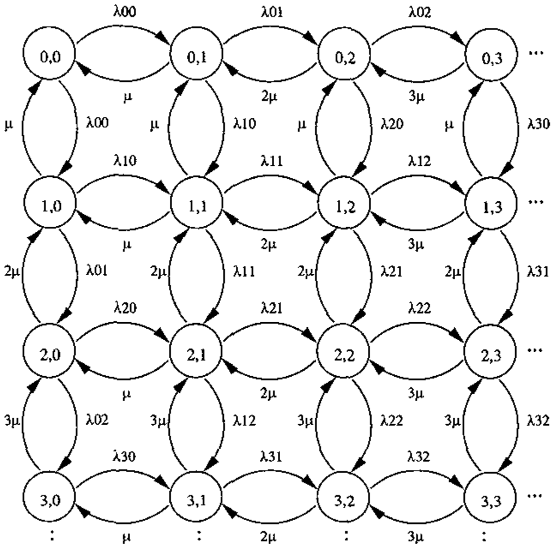

Fig. 6. Markov chain for two band system.

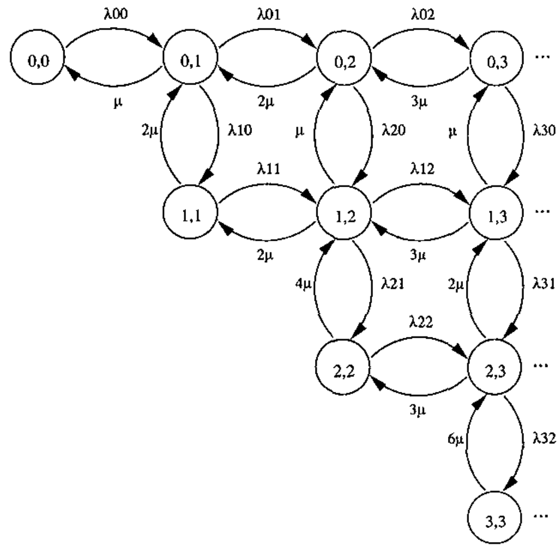

Fig. 7. Collapsed Markov chain for two band system.

sional (or N dimensional for N bands). For simplicity, this paper will only treat the case of hard handoff.

One obvious strategy is to assign a call to a band randomly. If the call is blocked in the chosen band, it is tried in the next band. This gives rise to the Markov chain of Figure 6, where λ_{ij} is the arrival rate to the band in state j from either state (i,j) or (j,i). The rate of arrivals to band k from a given state is the overall arrival rate ($= \lambda$) times the probability of initially selecting band k ($= 1/2$), times the probability of not being blocked on band k given that band k is in state j ($= 1 - B_j$), plus the overall arrival rate times the probability of initially selecting the *other* band ($= 1/2$), but being blocked on that band ($= B_i$) and not being blocked on band k ($= 1 - B_j$). Thus $\lambda_{ij} = (\lambda/2)((1-B_j)+B_i(1-B_j)) = \lambda(1-B_j)(1+B_i)/2$. By symmetry, this collapses to the chain of Figure 7, but now with λ_{ii} doubled to $\lambda(1 - B_i^2)$. That is because here the probability of being in state (i,i) is the probability of having i calls in each of the two bands, while the probability of being in state (i,j), $j \neq i$, is the probability of *either* having i calls in the first band and j in the second, *or* having j calls in the first band and i in the second. Note that this does not require knowledge of the blocking probability for each state, but only the marginal blocking probability given there are i calls in a particular band.

Another strategy which has proven successful is to assign new calls to the band which currently has the fewest calls [5], with blocked calls again trying the other band. This can again be represented by the collapsed Markov chain of Figure 7, but now with $\lambda_{ij} = \lambda(1 - B_j)$ if $i > j$, $\lambda_{ij} = \lambda B_i(1 - B_j)$ if $i < j$, and $\lambda_{ii} = \lambda(1 - B_i^2)$. This "least load" ("LL") strategy has been analysed thoroughly under the assumption of no blocking [6].

Figures 8 and 9 show the actual (i.e., simulated) and Markov state distributions for each of the two strategies, along with the M/M/∞ results for comparison. The spreading factor was halved to 64, so that the same total bandwidth was used as for the single band case, and the offered load was 17 Erlangs. For simplicity, the graph shows the marginal probability of having i calls in the first band, rather than the full distribution. Clearly there is very good agreement between the model and simulation. Least load allocation shows a more peaked distribution, while random allocation shows a heavier tail, with more cells having a large number of calls. This behaviour is captured well by the proposed Markov model.

The blocking probabilities for the simulated (respectively Markov) cases were 16.1% (15.6%) for least load, and 16.6% (13.4%) for random. The figures for the Markov model are an approximation, given by the square of the marginal per-band blocking probability. This is a fair approximation because a call must be rejected on both bands to be blocked.

To show once again that the good agreement with simulation is not solely due to the use of accurate per-state blocking probabilities, Figures 10 and 11 show the Markov distribution for each strategy using the per-state blocking probabilities from the *other* strategy. Each graph shows the two sets of simulation results and the results using one of the Markov chains with both sets of measured blocking probabilities. In each case, the predicted results match the

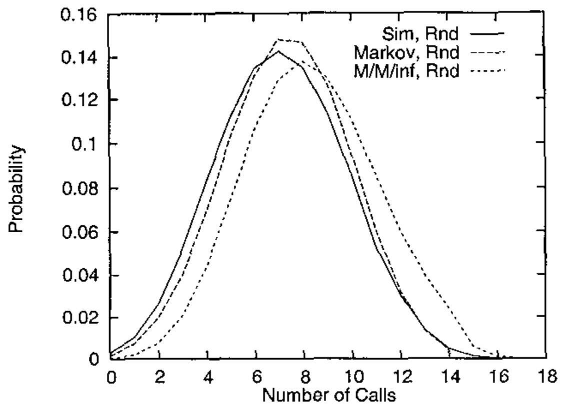

Fig. 8. Simulated vs. Markov and M/M/∞ results for random band allocation strategy.

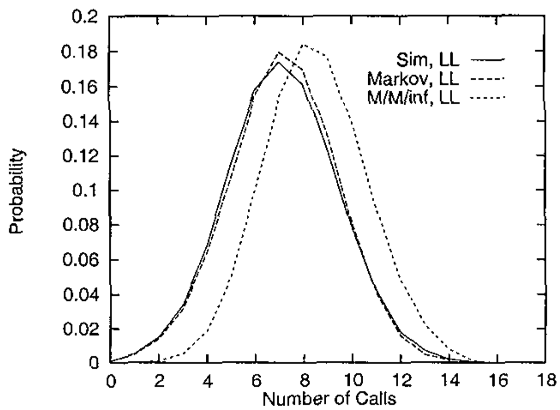

Fig. 9. Simulated vs. Markov and M/M/∞ results for least load band allocation strategy.

simulation to which the Markov model corresponds more closely than the one from which the per-state blocking probabilities were taken. This confirms that the model is not overly sensitive to the exact per-state blocking probabilities. Thus this model can be used to predict the overall blocking probability for different band allocation schemes based on the different Markov chains and a single set of measured per-state blocking probabilities.

V. Conclusion

It is possible to model soft blocking in multi-cell CDMA systems as an independent birth and death process at each cell. The birth process is thinned by the state-dependent blocking probability. The birth process can also be modified to model soft handoff, and the process can be extended to analyse multi-band systems.

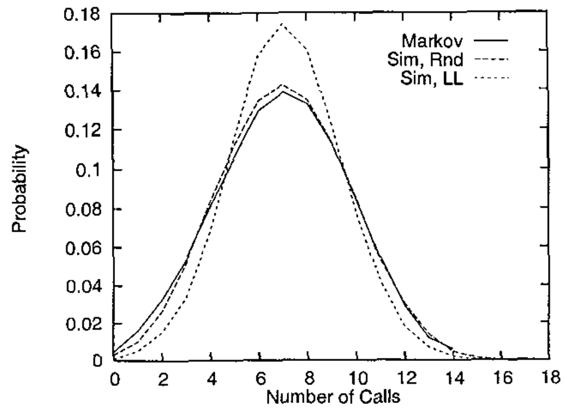

Fig. 10. Predicted state distributions for random Markov model using least load blocking probabilities.

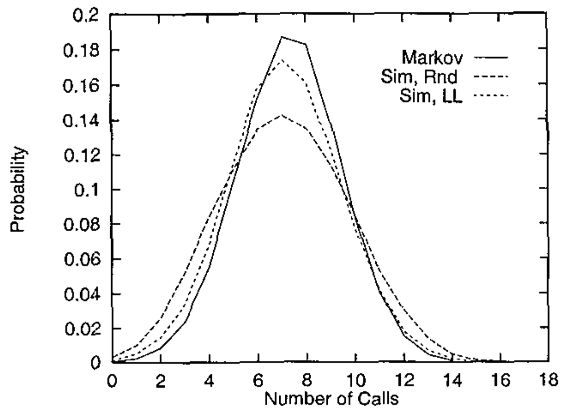

Fig. 11. Predicted state distributions for least load Markov model using random blocking probabilities.

Acknowledgement

The authors thank Chaitanya Rao for comments on this manuscript, and for pointing out an error in the original arrival rates for the multiband Markov chain with random assignment.

References

[1] A. J. Viterbi, *Principles of Spread Spectrum Communication*. Reading, MA: Addison-Wesley, 1995.

[2] Y. Ishikawa and N. Umeda, "Capacity design and performance of call admission control in cellular CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1627–1635, Oct. 1997.

[3] M. Kijima, *Markov processes for stochastic modeling*. London: Chapman and Hall, 1997.

[4] L. L. H. Andrew, "Measurement-based band allocation in multiband CDMA," in *Proc. Infocom '99*, (New York), 1999, pp. 1536–1543.

[5] T. Dean, P. Fleming, and A. Stolyar, "Estimates of multicarrier CDMA system capacity," in *Proc. Winter Sim. Conf.*, (Washington, DC), 1998, pp. 1615–1622.

[6] P. J. Fleming and B. Simon, "Heavy traffic approximations for a system of infinite servers with load balancing," to appear.