Performability modelling of wireless communication systems

Kishor S. Trivedi¹,†,‡, Xiaomin Ma²,§ and S. Dharmaraja³,*,¶

¹CACC, Department of ECE, Duke University, Durham, NC 27708
²Engineering & Physics Department, Oral Roberts University, Tulsa, OK 74171
³TRLabs, 10-75 Scurfield Boulevard, Winnipeg, MB R3Y 1P6, Canada

SUMMARY

The high expectations of performance and availability for wireless mobile systems has presented great challenges in the modelling and design of fault tolerant wireless systems. The proper modelling methodology to study the degradation of such systems is so-called performability modelling. In this paper, we give overview of approaches for the construction and the solution of performability models for wireless cellular networks. First, we start with the Erlang loss model, in which hierarchical and composite Markov chains are constructed to obtain loss formulas for a system with channel failures. Consequently, we develop two level hierarchical models for the wireless cellular system with handoff and channel failures. Then, for a TDMA system consisting of base repeaters and a control channel, we build a hierarchical Markov chain model for automatic protection switching (APS). Finally, we discuss stochastic reward net (SRN) models for performability analysis of wireless systems. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: channel allocation; handoff; Markov chain; Markov reward model; performability; stochastic reward net; wireless communication systems

1. INTRODUCTION

With the rapid growth of wireless communication services, customers are expecting the same level of availability and performance from wireless communication systems as traditional wireline networks. In general, wireless systems are characterized by their scarce radio resources which limit not only the service offering but also the quality of service (QoS). Furthermore, service degradation can also be caused by component failures, software failures and human-errors in operation in the wireless system. A failure, depending on its nature, may have different impacts on system performance. The high degree of mobility enjoyed in wireless networks is in turn the cause of inherent unreliability. Compared with wired networks, wireless networks need

*Correspondence to: S. Dharmaraja, TRLabs, 10-75 Scurfield Boulevard, Winnipeg, MB R3Y 1P6, Canada.
†This work was done while K. Trivedi was a visiting Professor in the Department of Computer Science and Engineering holding the Poonam and Prabhu Goel Chair at the Indian Institute of Technology, Kanpur.
‡E-mail: kst@ee.duke.edu
§E-mail: xma@oru.edu
¶E-mail: dharmar@win.trlabs.ca

Contract/grant sponsor: AFOSR MURI; contract/grant number: F49620-1-0327

Published online 12 May 2003
Copyright © 2003 John Wiley & Sons, Ltd.
to deal with disconnects due to handoff [1,2], noise and interference, fast (slow) fading, blocked and weak signals and run-down batteries [3,4], etc. In addition, the performance and availability of a wireless system is affected by the outage-and-recovery of its supporting functional units. From the designer and operator’s point of view, it is of great importance to take these factors into account integratively.

Traditional pure performance model that ignores failure and recovery but considers resource contention generally overestimates the system’s ability to perform. On the other hand, pure availability analysis tends to be too conservative since performance considerations are not taken into account. To obtain realistic composite performance and availability measures, one should consider performance changes that are associated with failure and recovery behaviour. The proper modelling methodology to study the degradation behaviour of the system is the so-called performability modelling [5–8], which takes into account both availability and performance, thus providing a more complete picture.

Generally, performability evaluation involves two steps: the construction of a suitable model and the solution of the model. Two common techniques of analysing performability are (i) a composite continuous-time Markov chain (CTMC) containing both the performance as well as the availability related events, (ii) a two-level hierarchical model where the upper level is a Markov reward model (MRM) that is essentially the availability model with each state of the MRM being assigned a reward rate derived from the lower-level pure performance model. One measure of performability can then be expressed as the expected steady-state reward rate:

$$E[Z] = \sum_j r_j \pi_j$$

where $r_j$ is the reward rate assigned to state $j$ and $\pi_j$ is the steady-state probability of state $j$ of the upper level MRM. If transient analysis is of interest, let $Z(t)$ be the reward rate at time $t$, then the expected reward rate at time $t$:

$$E[Z(t)] = \sum_j r_j \pi_j(t)$$

For an irreducible CTMC, the expected steady-state reward rate is also expressed as

$$\lim_{t \to \infty} E[Z(t)] = E[Z] = \sum_j r_j \pi_j$$

The expected accumulated reward in the interval $(0, t]$ can be computed:

$$E[Y(t)] = \sum_j r_j \int_0^t \pi_j(\tau) d\tau = \sum_j r_j L_j(t)$$

where $Y(t) = \int_0^t Z(\tau) d\tau$, $\pi_j(t)$ is the transient probability of the upper level MRM being in state $j$ at time $t$, and $L_j(t)$ is the expected total time spent by the upper level MRM in state $j$ during $(0, t]$.

In this paper, we report recent approaches to performability modelling of wireless communication networks. The paper is organized as follows. In Section 2, we construct composite and hierarchical CTMC models for the combined performance and availability analysis of general wireless systems. In Section 3, the performability of wireless cellular systems with handoff is considered and analysed. In Section 4, we build hierarchical Markov chain models for APS in TDMA system consisting of base repeaters and a control channel. In Section 5, we demonstrate several stochastic reward net (SRN) models for the performability analysis of wireless systems. Finally, we make our conclusions in Section 6.
2. ERLANG LOSS MODEL [7]

Consider a telephone switching system consisting of \( n \) trunks (or channels) with an infinite caller population. If an arriving call finds all \( n \) trunks busy, it does not enter the system and is lost instead. The call arrival process is assumed to be Poissonian with rate \( \lambda \). We assume that call holding times are independent, exponentially distributed random variables with parameter \( \mu \) and independent of the call arrival process. Assume that the times to trunk failure and repair are exponentially distributed with mean \( \frac{1}{\gamma} \) and \( \frac{1}{\tau} \), respectively. Also assume that a single repair facility is shared by all the trunks.

We first construct the composite model for the combined performance and availability analysis and the state diagram is shown in Figure 1. Here, the state \((i,j)\) denotes \( i \) non-failed trunks and \( j \leq i \) ongoing calls in the system. Note that trunks that are in use as well as those that are free can fail with the corresponding failure rate. This composite model is a homogeneous irreducible CTMC with \((n + 1)(n + 2)/2\) states, and the steady-state probability can be obtained by solving the linear system of homogeneous equations. Such a solution may be obtained using a software package such as SHARPE [9]. The total call blocking probability is then given by

\[
T_b = \sum_{i=0}^{n} \pi_{i,i}
\]

Figure 1. State diagram for the Erlang loss composite model.
The above composite performability model might encounter two problems: largeness and stiffness. Largeness means that finding the required measures will be cumbersome and numerically error-prone when the number of trunks is large. On the other hand, stiffness means that transition rates in the Markov model range over many orders of magnitude. To avoid the problems of largeness and stiffness, we can compute the required measure approximately using a hierarchical approach [4,8,10,11]. In this approach, a top-level availability model (Figure 2) is turned into a Markov reward model (MRM), where the reward rates come from a sequence of performance models (Figure 3) and are supplied to the top-level availability model.

We first present an availability model that accounts for failure-repair behaviour of trunks (we will use trunks and channels interchangeably in this paper); then, we use a performance model to compute performance indices such as blocking probability given the number of non-failed trunks; finally, we combine the two models together and give performability measures of interest. The availability model is then a homogeneous CTMC with the state diagram shown in Figure 2. Here, the state index denotes the number of non-failed trunks in the system. The steady-state probability for the number of non-failed channels in the system is given by

$$p_i = \frac{1}{i!}(\tau/\gamma)^i p_0, \quad i = 1, 2, \ldots, n$$

where the steady-state system unavailability:

$$U = p_0 = \left[ \sum_{i=0}^{n} \frac{1}{i!}(\tau/\gamma)^i \right]^{-1}$$

Consider the performance model with the given number $i$ of non-failed channels. The quantity of interest is the blocking probability, that is, the steady-state probability that all trunks are busy, in which case the arriving call is refused service. Note that in this performance model, the assumption is that blocked calls are lost (not re-attempted). The performance model of this telephony system is an $M/M/i$ loss system, and the state diagram is shown in Figure 3. The blocking probability with $i$ channels in the system is given by

$$P_b(i) = \frac{(\lambda/\mu)^i/i!}{\sum_{j=0}^{i} (\lambda/\mu)^j/j!}$$
This equation is known as the *Erlang’s B loss formula*. It can be shown to hold even if the call holding time follows arbitrary distribution with mean $1/\mu$ [12].

Attach a reward rate $r_i$ to the state $i$ of the availability model as the blocking probability with $i$ trunks in the system, that is, $r_i = P_b(i)$, $i \geq 1$ and $r_0 = 1$. Then the required total blocking probability can be computed as the expected reward rate in the steady-state and is given by

$$
\hat{T}_b = \sum_{i=0}^{n} r_i \pi_i = \pi_0 + P_b(n)\pi_n + \left[ \sum_{i=1}^{n-1} P_b(i)\pi_i \right] 
$$

(1)

where $\pi_i$ is the steady-state probability that $i$ non-failed trunks are there in the system.

The total loss probability expression above can be seen to consist of three summands: the first part is system unavailability $U$, the second part is the call blocking probability due to buffer full weighted by the probability that the system is up and the bracketed part on the right-hand side of Equation (1) is the buffer full probability in each of the degraded states weighted by the probability of the corresponding degraded state.

In Figure 4, we compare the exact total blocking probability $T_b$ with approximate result $\hat{T}_b$ as functions of the number of trunks. The error incurred by the two-level performability model is negligible in this case. This will normally be the case when the performance-related events are relatively fast (by a few orders of magnitude) when compared with failure-related events.

### 3. MODELLING CELLULAR SYSTEMS WITH FAILURE AND REPAIR [7]

The above Erlang loss formula cannot be used in cellular wireless networks due to the phenomenon of handoff. In this section, we discuss a two level hierarchical performability model for wireless cellular networks with handoff.
The object under study is a typical cellular wireless system. In the system, mobile subscribers (MSs) are provided with telephone service within a geographical area. The service area is divided into multiple adjacent cells. MSs communicate via radio links to base stations (BSs), one for each cell. When an MS moves across a cell boundary, the channel in the old BS is released and an idle channel is required in the new BS. This phenomenon is called handoff. Handoff is an important function of mobility management. To reduce the dropping probability of handoff calls, a fixed number of guard channels is reserved exclusively for the handoff calls [13].

We consider a single cell with limited number of channels, \( n \), in the channel pool. Let the number of guard channels, \( g (g < n) \), be reserved exclusively for handoff calls. Changing the number of guard channels results in different new call blocking probability and handoff call dropping probability. We notice that channel failures are rather rare compared with the new call or handoff call arrivals and departures. Consequently, we use hierarchical decomposition to obtain an approximate solution: we first present an upper level availability model which accounts for the possible channel failures and repairs. Then, we compute performance indices by constructing a lower level CTMC performance model. Finally, we combine them together and give performability measures of interest.

The upper level model, as shown in Figure 2, describing the failure and repair behaviour of the system, is a pure availability model. Let \( \pi_i \) \((i \in \{0, 1, 2, \ldots, n\})\) be the steady-state probability of the CTMC being in state \( i \) of the upper level model. We know that

\[
\pi_i = \frac{1}{i!} \left( \frac{\tau}{\gamma} \right)^i \pi_0
\]

where \( \pi_0 \) is the steady-state probability of the CTMC in state 0 which equals the steady state unavailability.

\[
\pi_0 = \left[ \sum_{i=0}^{n} \frac{1}{i!} \left( \frac{\tau}{\gamma} \right)^i \right]^{-1} = U
\]

The lower level model, as shown in Figure 5, captures the pure performance aspect of the system [14]. Each state represents the number of talking channels in the system. In Figure 5, \( i \in \{1, 2, \ldots, n\} \). Let \( \lambda_1 \) be the rate of Poisson arrival stream of new calls and \( \lambda_2 \) be the rate of Poisson stream of handoff arrivals. Let \( \mu_1 \) be the rate that an ongoing call (new or handoff) completes service and \( \mu_2 \) be the rate at which the mobile engaged in the call departs the cell. When an idle channel is available in the channel pool and a handoff call arrives, the call is accepted and a channel is assigned to it. Otherwise, the handoff call is dropped. When a new call arrives, it is accepted provided that at least \( g+1 \) idle channels are available in the channel pool; otherwise, the new call is blocked.

Figure 5. CTMC Performance model of wireless handoff.
The state-dependent arrival and departure rates in the birth-death process of Figure 5 are given by

\[ \Lambda(j) = \begin{cases} \lambda_1 + \lambda_2, & \text{if } j < i - g \\ \lambda_2, & \text{if } i - g < j < i \end{cases} \] (4)

and \( \mu(j) = j(\mu_1 + \mu_2), j = 1, 2, \ldots, i. \)

Let the steady-state probability of the CTMC being in state \( j \) be denoted by \( \pi_j \).

Let \( l = \lambda_1 + \lambda_2 \), \( \mu = \mu_1 + \mu_2 \), and \( A = \lambda/\mu \), \( A_1 = \lambda_2/(\mu_1 + \mu_2) \).

After finding \( \pi_j \) \[14\], we can write the expression for the dropping probability for handoff calls \( P_d(i) \) in the lower model.

\[ P_d(i) = \frac{A^{i-g} \mu_{1-g}}{i!} A_1^{i-g} \prod_{j=0}^{i-g-1} \frac{A_j}{j!} \] (5)

Similarly, the expression for the blocking probability of new calls \( P_b(i) \) in the lower level model:

\[ P_b(i) = \sum_{j=1}^{i} \pi_j = \frac{\sum_{j=i-g}^{i} A^{j-i} \mu_{j-i} A_1^{j-i-g} \prod_{j=0}^{i-j-g-1} \frac{A_j}{j!}}{\sum_{j=0}^{i-g-1} \frac{A_j}{j!} + \sum_{j=i-g}^{i} \frac{A^{j-i} \mu_{j-i} A_1^{j-i-g} \prod_{j=0}^{i-j-g-1} \frac{A_j}{j!}}{j!} A_1^{j-i-g}} \] (6)

For fast stable computation of above dropping probability and blocking probability, the optimization problems to determine the optimal number of channels, and the fixed-point iteration-based scheme to determine handoff arrival rate, we refer readers to Reference \[14\].

To get the numerical measures for the whole system, the lower level performance model is solved and its results are passed as reward rates to the upper level availability model.

We denote, respectively, \( P_d(i) \) and \( P_b(i) \) as the total dropping and total blocking probability obtained from the performability model. The approximate dropping probability is obtained as

\[ P_d = U + P_d(n) \cdot \pi_n + \sum_{i=1}^{n-1} (\pi_i \cdot P_d(i)) \] (7)

The approximate blocking probability is obtained as

\[ P_b = U + \sum_{i=1}^{n} \pi_i + (\pi_n \cdot P_b(n)) + \sum_{i=g+1}^{n-1} (\pi_i \cdot P_b(i)) \] (8)

Similarly, the above approximate loss probabilities consist of unavailability and performance loss due to channel resource full as well as degraded buffer full.

4. HIERARCHICAL MODEL FOR APS IN A TDMA SYSTEM \[15\]

4.1. Wireless cellular systems with failures

A TDMA system with hard handoff in which a cell has multiple base repeaters, say \( N_b \), is considered in this model. Each base repeater provides a number of channels, say \( M \), for mobile
terminals to communicate with the system. Therefore a total of $N_bM$ channels are available when the whole system is working properly. Normally, one of the channels is dedicated to transmitting control channel messages. Such a channel is called a control channel. The total number of available voice channels is then $N_bM - 1$. We also assume that the control channel is selected randomly out of $N_bM$ channels. Failure of the control channel will cause the whole system to fail. To avoid this undesirable situation, an automatic protection switching (APS) scheme is suggested in Reference [16] so that the system automatically selects a channel from the rest of the non-failed channels to substitute the failed control channel. If all non-failed channels are in use (talking), then one of them is forcefully terminated and is used as the control channel.

A cell as a whole is subject to failures which make all channels in it inaccessible, causing a full outage. In practice, this type of failure may occur when the communication links between base station controller and base repeaters do not function properly, or a critical function unit (such as base station controller) fails. In this model, we will refer to this type of failure as the platform failure. Each base repeater is also subject to failure which disables the channels that it provides. In a system without APS, if a failed base repeater happens to be the one hosting the control channel, it results in a full outage, same as the situation caused by a platform failure.

We use the traditional two-level performability model: we first present an availability model which accounts for the failure and repair of base repeaters; second, we use a performance model to compute performance indices given the number of non-failed base repeaters; finally, we combine them together and give corresponding loss formulas.

### 4.2. The availability model

All failure events are assumed to be mutually independent. Times to platform failure and repair are assumed to be exponentially distributed with mean $1/\lambda_s$ and $1/\mu_s$, respectively. Also assume that times to base repeater failure and repair are exponentially distributed with mean $1/\lambda_b$ and $1/\mu_b$, respectively, and that a single repair facility is shared by all the base repeaters.

Let $s \in S = \{0, 1\}$ denote a binary value indicating whether or not the system is down due to a platform failure (0: system down due to a platform failure; 1: no platform failure has occurred). Also let $k \in B = \{0, 1, \ldots, N_b\}$ denote the number of non-failed base repeaters. The 2-tuple $(s, k), s \in S, k \in B$ defines a state in which the system is undergoing a (no) platform failure if $s = 0$ (if $s = 1$) and $k$ base repeaters are up. The underlying stochastic process is a homogeneous CTMC with state space $S \times B$. Let $\pi(s, k; N_b)$ be the corresponding steady-state probability. The state diagram of this irreducible CTMC is depicted in Figure 6.

![Figure 6. Markov chain of availability model.](image-url)
Solving the CTMC, we have

\[
\pi(s, k; N_b) = \begin{cases} 
\frac{\lambda_s}{\lambda_s + \mu_s} \frac{1}{k!} \left( \frac{\mu_b}{\lambda_b} \right)^k \left[ 1 + \sum_{j=1}^{N_b} \frac{1}{j!} \left( \frac{\mu_b}{\lambda_b} \right)^j \right]^{-1}, & \text{if } s = 0 \\
\frac{\mu_s}{\lambda_s + \mu_s} \frac{1}{k!} \left( \frac{\mu_b}{\lambda_b} \right)^k \left[ 1 + \sum_{j=1}^{N_b} \frac{1}{j!} \left( \frac{\mu_b}{\lambda_b} \right)^j \right]^{-1}, & \text{if } s = 1
\end{cases}
\]  

The system unavailability corresponds to all the states in which either the system has a platform failure that brings the whole system down, or in a system without APS, a base repeater hosting the control channel fails, or the system even without platform failure has no working base repeaters happens to host the control channel. For a system without APS, the probability that one of the \((N_b - k)\) failed base repeaters happens to host the control channel is \(N_b / C_0\). Denote \(U(N_b)\) as the steady-state system unavailability. For both systems with and without APS, we thus write unavailability as

\[
U(N_b) = \begin{cases} 
\sum_{k=0}^{N_b} \pi(0, k; N_b) + \sum_{k=0}^{N_b} \pi(1, k; N_b) \frac{N_b - k}{N_b}, & \text{w/o APS} \\
\sum_{k=0}^{N_b} \pi(0, k; N_b) + \pi(1, 0; N_b), & \text{w/ APS}
\end{cases}
\]  

4.3. Performability

For each of the states of the availability model of Figure 6, Equations (5) and (6) in Section 3 provide performance indices given the number of non-failed channels.

We notice that calls can be blocked (or dropped) due to system being down or being full. The former type of loss is captured by the pure availability model while the latter type of loss is captured by the pure performance model. We now wish to combine the two types of losses. The primary vehicle for doing this is to determine pure performance losses for each of the availability model states and attach these loss probabilities as reward rates (or weights) to these states. Such a Markov reward model has been called a performability model. We list reward rates for the states of the availability model in Table I for systems without APS and Table II for system with APS. We first consider system states that are down states.

**Table I. Reward rates for systems without APS.**

<table>
<thead>
<tr>
<th>State ((s, k))</th>
<th>New call blocking</th>
<th>Handoff call dropping</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0, k), k = 0, \ldots, N_b)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>((1, 0))</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>((1, k), k = 1, \ldots, N_b)</td>
<td>(\frac{N_b - k}{N_b} + P_b^{(i)}(kM - 1)\frac{k}{N_b}), (\text{o.w.})</td>
<td>(\frac{N_b - k}{N_b} + P_d^{(i)}(kM - 1)\frac{k}{N_b})</td>
</tr>
</tbody>
</table>

Copyright © 2003 John Wiley & Sons, Ltd.  
Clearly, for both systems without and with APS, a cell is not able to accept any new calls or handoff calls if it has platform failure which corresponds to the states $(0, k)$ for $k = 0, \ldots, N_b$, or all base repeaters are down which corresponds to the state $(1, 0)$. Therefore, reward rates of both overall new call blocking and handoff call dropping are 1’s.

In addition, for a system without APS, control channel may go down in states $(1, k)$ for $k = 1, \ldots, N_b$ with probability $(N_b - k)/N_b$ and cause new call blocking and handoff call dropping. This corresponds to the rates with $(N_b - k)/N_b$ in the last row of Table I. All cases mentioned above contribute to system unavailability, $U(N_b)$. Hence, system unavailability, $U(N_b)$, also consists of one of the parts of the overall new call blocking probability and handoff call dropping probability.

We now consider states in which the system is not undergoing a full outage caused by failures of platform, control channel (if system w/o APS) or all base repeaters being down.

The corresponding states are $(1, k)$ for $k = 1, \ldots, N_b$. The total number of non-failed channels in state $(1, k)$ is $kM - 1$. Thus, new call blocking probability and handoff call dropping probability in these states are $P_b^{(k)}(kM - 1)$ and $P_d^{(k)}(kM - 1)$, respectively. Thus, these probabilities are used as reward rates to these states for overall new call blocking and handoff call dropping.

For a system without APS, we note that the probability of not having the control channel down in state $(1, k)$ for $k > 0$ is $k/N_b$. Therefore, the reward rates, $P_b^{(k)}(kM - 1)$ and $P_d^{(k)}(kM - 1)$, are also weighted by $k/N_b$ (shown in the last row of Table I).

Also, in case that the number of idle channels is less than the number of guard channels, i.e. $kM - 1 < g$ for states $(1, k)$, $k = 1, \ldots, N_b$, a cell is not able to set up any new calls because all available channels are reserved for handoff calls. Hence, the reward rates for new call blocking assigned to the corresponding states are 1’s.

Now let $G = \lfloor (g + 1)/M \rfloor$. Summarizing Tables I and II, the total call blocking probability can be written as the expected steady state reward rate,

$$P_b^{(k)}(N_b) = U(N_b) + \begin{cases} 
  I(G > 0) \sum_{k=1}^{G} \pi(1, k; N_b) \left( \frac{k}{N_b} \right) 
  + \sum_{k=G+1}^{N_b} \pi(1, k; N_b) P_b^{(k)}(kM - 1) \left( \frac{k}{N_b} \right), & \text{w/o APS} \\
  I(G > 0) \sum_{k=1}^{G} \pi(1, k; N_b) 
  + \sum_{k=G+1}^{N_b} \pi(1, k; N_b) P_b^{(k)}(kM - 1), & \text{w/o APS} 
\end{cases}$$

(11)

Copyright © 2003 John Wiley & Sons, Ltd.

where $I(e)$ is the indicator function: $I(e) = 1$ if expression $e$ is true; $I(e) = 0$, otherwise. Similarly the total handoff call dropping probability can be given as

$$P_d^{(i)}(N_b) = U(N_b) + \begin{cases} \sum_{k=1}^{N_b} \pi(1,k;N_b)P_d^{(i)}(kM - 1)\frac{k}{N_b}, & \text{w/o APS} \\ \sum_{k=1}^{N_b} \pi(1,k;N_b)P_d^{(i)}(kM - 1), & \text{w/ APS} \end{cases}$$ (12)

We should note that the hierarchical approach we have followed to obtain performability expressions is indeed an approximate solution to the model. Since we are interested in the

Table III. Parameters used in numerical study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_b$</td>
<td>Number of base repeaters</td>
<td>10</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of channels/base repeater</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>New-call arrival rate</td>
<td>20 calls/min</td>
</tr>
<tr>
<td>$1/\mu_1$</td>
<td>Mean call holding time</td>
<td>2.5 min</td>
</tr>
<tr>
<td>$1/\mu_2$</td>
<td>Mean time to handout</td>
<td>1.25 min</td>
</tr>
<tr>
<td>$\lambda_s$</td>
<td>Platform failure rate</td>
<td>1/year</td>
</tr>
<tr>
<td>$1/\mu_s$</td>
<td>Mean repair time of platform</td>
<td>8 h</td>
</tr>
<tr>
<td>$\lambda_b$</td>
<td>Base repeater failure rate</td>
<td>2/year</td>
</tr>
<tr>
<td>$1/\mu_b$</td>
<td>Mean repair time of base repeater</td>
<td>2 h</td>
</tr>
</tbody>
</table>

Figure 7. $P_b^{(i)}(N_b)$ versus $q$ for systems without APS and with APS (top); percentage of unavailability $U(N_b)$ in $P_b^{(i)}(N_b)$ (bottom).
steady-state performability measures rather than those in the transient regime, we neglect the fact that a failing base repeater may also bluntly discard all ongoing calls on it and therefore cause call dropping. We consider these simplifications to have a negligible effect on the steady-state measures.

We now present numerical results and Table III summarizes the parameters used. In Figures 7 and 8, for both systems without APS and with APS, we plot new-call blocking probability and handoff-call dropping probability, respectively, against new-call arrival rate, \( \lambda_1 \). The plots show that both probabilities increase but stay nearly flat when new call traffic is low (\(<20\) calls/min). The probabilities then increase sharply after \( \lambda_1 \) exceeds 20 calls/min. The improvement by APS can be seen as reductions of blocking probability and dropping probability. Improvement remains steady given low traffic but diminishes rapidly as traffic becomes heavier.

5. SRN MODELS FOR WIRELESS CELLULAR SYSTEMS WITH FAILURES

In order to automate the generation and solution of large CTMCs or MRMs, a higher-level language is often desired, stochastic Petri net (SPN) and its derivatives are commonly used for this purpose.
5.1. Introduction to SRN

Stochastic reward net (SRN) is an extension of Petri net (PN), which is a high level description language for formally specifying complex systems. A PN is a bipartite directed graph with two types of nodes: places and transitions. Each place may contain an arbitrary (natural) number of tokens. For a graphical presentation, places are depicted as circles, transitions are represented by bars and tokens are represented by dots or integers in the places. Each transition may have zero or more input arcs, coming from its input places; and zero or more output arcs, going to its output places. A transition is enabled if all of its input places have at least as many tokens as the multiplicity of the corresponding input arc. When enabled, a transition can fire and will remove from each input place and add to each output place the number of tokens corresponding to the multiplicities of the input/output arcs. A marking depicts the state of a PN which is characterized by the assignment of tokens to all its places. With respect to a given initial markings, its reachability set is defined as the set of all markings that are reachable by means of a firing sequence of transitions starting from the initial marking. To get the performance and reliability/availability measures of a system, appropriate reward rates are assigned to its SRN. As SRN is automatically transformed into a Markov reward model, steady state and/or transient analysis of the Markov reward model produces the required measures of the original SRN. Once the SRN is formulated a software package such as SPNP [17] or the latest version of SHARPE can be used to specify and solve the SRN model.

5.2. Basic hierarchical performability model

Here, we consider the same wireless cellular system as the system in Section 3. We build a two level SRN model for performability analysis. The upper level model, as shown in Figure 9, describes the failure and repair behaviour of the system. The number of tokens in place $T$ represents the number of channels that are currently non-failed in the cell. The number of tokens in place $R$ represents the number of channels that have failed. Transition $Tr$ with rate $\tau$ represents the repair of a channel while transition $Tf$ with label $\gamma$ represents the failure of a channel. The actual firing rate of $Tf$ equals the number of tokens in place $T$ multiplied by $\gamma$; this is indicated by the ‘#’ next the arc from place $T$ to transition $Tf$.

The lower level SRN model, as shown in Figure 10, captures the pure performance aspect of the system. In Figure 10, $i \in \{1, 2, \ldots, n\}$. The number of tokens in place $P_{\text{talk}}$ represents the

![Figure 9. Upper level pure availability model.](image-url)
The firing of transition $T_{\text{new\_call}}$ represents the arrival of a new call and the firing of transition $T_{\text{handoff\_in}}$ represents the arrival of a handoff call from one of the neighbouring cells. A handoff call will be dropped only when all channels are occupied (i.e. $\#P_{\text{talk}} = i$). This is realized by an inhibitor arc from place $P_{\text{talk}}$ to $T_{\text{handoff\_in}}$ with multiplicity $i$. A new call, however, will be blocked if there are no more than $g$ idle channels. This is simply reflected in the SRN by the inhibitor arc from place $P_{\text{talk}}$ to transition $T_{\text{new\_call}}$ with multiplicity $i - g$. The firings of transition $T_{\text{call\_completion}}$ and $T_{\text{handoff\_out}}$ represent the completion of a call and the departure of an outgoing handoff call, respectively. The rates of transitions $T_{\text{call\_completion}}$ and $T_{\text{handoff\_out}}$ are marking-dependent, as indicated by the two “#” symbols next to the arcs from the place $P_{\text{talk}}$.

Two steady-state measures are of interest from the SRN model of Figure 10, namely, the new-call blocking probability, $P_{\text{b}}^{(i)}(i)$, and the handoff-call dropping probability, $P_{\text{d}}^{(i)}(i)$. We obtain these two measures by computing the expected steady-state reward rate for the SRN model with the proper assignment of reward rates to the markings.

\[
P_{\text{b}}^{(i)}(i) = \sum_{j \in \Omega} (r_{b})_j \pi^{(i)}_j \tag{13}
\]

\[
P_{\text{d}}^{(i)}(i) = \sum_{j \in \Omega} (r_{d})_j \pi^{(i)}_j \tag{14}
\]

The reward rate assigned to marking $j$ for computing the new-call blocking probability is

\[
(r_{b})_j = \begin{cases} 1, & \#P_{\text{talk}} \geq i - g \\ 0, & \#P_{\text{talk}} < i - g \end{cases}
\]

and that for the handoff dropping probability is

\[
(r_{d})_j = \begin{cases} 1, & \#P_{\text{talk}} = i \\ 0, & \#P_{\text{talk}} < i \end{cases}
\]

Thus, we denote respectively $P_{\text{d}}^{(i)}$ and $P_{\text{b}}^{(i)}$ as the total dropping and total blocking probability obtained from the performability model. This approximate total dropping probability is then obtained as

\[
P_{\text{d}}^{(i)} = \pi_0 + \pi_n \cdot P_{\text{d}}^{(i)}(n) + \sum_{i=1}^{n-1} (\pi_i \cdot P_{\text{d}}^{(i)}(i)) \tag{15}
\]
where $\pi_i (i \in \{0, 1, 2, \ldots, n\})$ is the steady-state probability of marking $i$ for the upper level model.

The approximate total blocking probability is obtained as

$$P_b^{(i)} = \pi_0 + \sum_{i=1}^{n} \pi_i + \pi_n \cdot P_b^{(i)}(n) + \sum_{i=g+1}^{n-1} (\pi_i \cdot P_b^{(i)}(i))$$

(16)

6. CONCLUSION

During the last decade we have witnessed a tremendous growth within the wireless communication industry. Customers want speed and improved performance, but only if it comes with reliable services. This requires fundamental rethinking of the traditional pure performance model that ignores failure, repair or recovery but mainly concentrates on resource contention. To reflect a real-world system in realistic way, availability, capacity and performance issues of a network should be considered in an integrated way.

In this paper, we have presented the CTMC, MRM and SRN models for performability study of a variety of wireless systems. By solving the two-level models, we can compute performability measures, such as call blocking probability and handoff call dropping probability, for wireless systems and wireless cellular systems with handoff, base repeaters, and control channels. Compared with composite models, the more robust and less time-consuming hierarchical models are known to provide high accuracy. It is expected that the models presented in this paper will be useful in wireless networks design and operation.

Further work might include a performability study of multi-media wireless system with multiple control channels and corresponding fault-tolerant protection schemes, and performability study of differentiated QoS services, IP wireless mobile systems, and survivability of cellular systems.

ACKNOWLEDGEMENTS

This research was supported by an AFOSR MURI grant no. F49620-1-0327.

REFERENCES


AUTHORS’ BIOGRAPHIES

**Kishor S. Trivedi** holds the Hudson Chair in the Department of Electrical and Computer Engineering at Duke University, Durham, NC. He is the Duke-Site Director of an NSF Industry-University Cooperative Research Center between NC State University and Duke University for carrying out applied research in computing and communications. He has been on the Duke faculty since 1975. He is the author of a well known text entitled, *Probability and Statistics with Reliability, Queueing and Computer Science Applications*, published by Prentice-Hall; this text was reprinted as an Indian edition; a thoroughly revised second edition (including its Indian edition) of this book has been published by John Wiley. He has also published two other books entitled, *Performance and Reliability Analysis of Computer Systems*, published by Kluwer Academic Publishers and *Queueing Networks and Markov Chains*, John Wiley. His research interests are in reliability and performance assessment of computer and communication systems. He has published over 300 articles and lectured extensively on these topics. He has supervised 36 PhD dissertations. He is a Fellow of the Institute of Electrical and Electronics Engineers. He is a Golden Core Member of IEEE Computer Society.

Kishor is spending a sabbatical year beginning Fall 2002 at IIT Kanpur where he is Poonam and Prabhu Goel Professor in the Department of Computer Science and Engineering. Kishor is also be a Fulbright Visiting Lecturer.

**Xiaomin Ma** received BE and ME degrees in electrical engineering in 1984 and 1989, respectively. He got the PhD degree in Information engineering at the Beijing University of Posts & Telecommunications, China, in 1999. From 2000 to 2002, he was a post-doctoral fellow at the Department of Electrical and Computer Engineering, Duke University, U.S.A. Currently, he is an assistant professor in the Engineering and Physics Department at Oral Roberts University in U.S. His research interests include stochastic modelling and analysis of computer and communication systems, computational intelligence and its applications to coding, signal processing, and control, and Quality of service (QoS) and Call admission control protocols in computer and communication networks.
S. Dharmaraja received the MSc degree in Applied Mathematics from Anna University, Madras, India, in 1994 and the PhD degree in Mathematics from the Indian Institute of Technology, Madras, in 1999. From 1999 to 2002, he was a postdoctoral fellow at the Department of Electrical and Computer Engineering, Duke University, U.S.A. Currently, he is a research associate at the TRLabs, Winnipeg, Canada. His research interests include applied probability, queuing theory, stochastic modelling and analysis of computer and communication systems.