Channel Allocation with Recovery Strategy in Wireless Networks

YUE MA, JAMES J. HAN
Global Software Division, Motorola, Inc., 1303 East Algonquin Road, IL01/ANX2, Schaumburg, IL 60196, USA
{a14438,cjh048}@email.mot.com

KISHOR S. TRIVEDI
Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA
kst@ee.duke.edu

Abstract. With the increasing penetration of wireless communications systems, customers are expecting the same level of service, reliability and performance from the wireless communication systems as the traditional wire-line networks. Due to the dynamic environment, such as the roaming of mobile subscribers, maintaining a high radio frequency (RF) availability is one of the most challenging aspects in wireless network management. To date, in wireless network management, a call is dropped when the channel it uses goes down. In order to increase system end-to-end availability, an RF channel recovery scheme is proposed in this paper. When an RF channel fails, the channel is replaced by another working channel and the call continues. The methods to replace failed RF channels of ongoing calls and to handle channel failures of handoff and new calls are investigated. Markov reward models are developed to compare system availability and performance. Automated generation and solution of Markov reward models is facilitated by a version of stochastic Petri nets that we call stochastic reward nets. The results show that the recovery scheme reduces the dropped calls and the blocked calls significantly under both light and normal traffics.

1 INTRODUCTION

With the increasing popularity of wireless communications systems, customers are expecting the same level of service, reliability and performance from the wireless communications systems as the traditional wire-line networks. Due to the dynamic environment, such as the roaming of mobile subscribers, maintaining a high radio frequency (RF) availability is one of the most challenging aspects in wireless networks. RF availability depends on natural environment, infrastructure, and subscriber handsets. There are many research and development results [1, 2] in each of the above areas. In this paper, our investigation focuses on RF channel failure recovery in the infrastructure, more specifically, in remote base site. There are many factors that cause RF failure, such as base repeater power failure, base repeater RF amplifier failure, etc. The main focus of this investigation is not on the causes of the RF failure. Instead, it is on the recovery method in the radio resource management of the base site, which maintains the ongoing calls.

In wireless networks, an RF channel is assigned to a call either during the call set-up process when a new call is initiated or during the handoff process when an ongoing call subscriber roams into the cell. Different channel allocation schemes have been studied [3, 4, 5, 6, 7, 8]. A common assumption in these studies has been that the channel in use never fails. However, in a practical environment, wireless networks, like any other physical system, are subject to failures. With the increasing penetration of wireless communications, a disruption in service could cause severe consequences in both economic and social sense. Thus providing restoration subsequent to channel failures has become an important issue in ensuring network integrity.

To obtain realistic performance measures for wireless networks, one should consider changes in performance due to failure related behavior. In performability analysis [9, 10], simultaneous consideration is given to both performance and reliability/availability measures. In this paper, an RF channel recovery scheme is proposed. When an RF channel fails, automatic repeat request (ARQ) will be first used for flow control. However, the ARQ schemes will not work in this situation since the channel is not in operation anymore. According to our recovery strategy, the failed channel is replaced by another working channel and the failed call continues. The methods to replace failed RF channels of ongoing calls and to handle channel failures of handoff and new calls are investigated.

With different new call arrival rates, the corresponding handoff arrival rates vary accordingly. To capture this dynamic behavior, a fixed-point iteration scheme [11] is applied to determine the handoff arrival rates. Basic modeling paradigm that we use is that of Markov reward models. In order to automatically generate and solve such models, we use the framework of stochastic reward nets. Stochastic reward net (SRN) models of RF channel allocation with and without channel recovery are developed to compare system availability and performance. SRN is an extension of Petri net [12, 13]. It has the advantage of specifying a real-world model in a compact and intuitive way. Since the early 90's, SRN has been used as a powerful modeling tool in performance, availability and reliability analysis in communications systems [14, 15, 16, 17, 18, 19]. In this paper, we applied stochastic Petri net package (SPNP) [20, 13] to calculate the numerical measures.

The paper is organized as follows. In Section 2, we give a brief introduction to SRNs. In Section 3, system specification is made for a generic single cell in a wireless network. In Section 4, a pure performance model for channel allocation is presented. In Section 5, we describe the fixed-point iteration scheme. In Section 6, we describe the channel recovery scheme in a progressive way. A simplified version is presented first, followed by a comprehensive one. In Section 7, we discuss how to express the numerical measures in terms of reward assignments in SRN. The numerical results for the performance model, one recovery scheme and one non-recovery method are compared in Section 8. Finally, we make our conclusions in Section 9.

2 INTRODUCTION TO STOCHASTIC REWARD NET (SRN)

Stochastic reward net [21] is an extension of Petri net (PN) [12, 13], 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 required by the multiplicities of the corresponding input arcs. 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 in all the places. With respect to a given *initial* marking, the *reachability set* is defined as the set of all markings reachable through any possible firing sequences of transitions, starting from the initial marking.

Generalized stochastic Petri nets (GSPNs) [12] extend the PNs by assigning a *firing time* to each transition. Transitions with exponentially distributed firing times are called *timed* transitions while the transitions with zero firing times are called *immediate* transitions. A marking in a GSPN is called *vanishing* if at least one immediate transition is enabled; otherwise it is called a *tangible* marking. For a given GSPN, an *extended reachability graph* ($\mathcal{ERG}$) is generated with the markings of the reachability set as the nodes and some stochastic information attached to the arcs, thus connecting the markings to each other. Under the condition that only a finite number of transitions can fire in finite time with non-zero probability, it can be shown that a given $\mathcal{ERG}$ can be reduced to a homogeneous continuous time Markov chain (CTMC) [12]. GSPN also introduces *inhibitor arcs*. An inhibitor arc from a place to a transition *disables* the transition if the place contains at least as many tokens as the cardinality of the inhibitor arc. Graphically, an inhibitor arc is represented by a line terminated with a small circle.

In order to make more compact models of complex systems, several extensions are made to GSPN, leading to the SRN. One of the most important features of SRN is its ability to allow extensive marking dependency. In an SRN, each tangible marking can be assigned with one or more *reward rate(s)*. Parameters such as the firing rate of the timed transitions, the multiplicities of input/output arcs and the reward rate in a marking can be specified as functions of the number of tokens in any place in the SRN. Another important characteristic of SRN is the ability to express complex enabling/disabling conditions through *guard* functions. This can greatly simplify the graphical representations of complex systems. For an SRN, all the output measures are expressed in terms of the expected values of the reward rate functions. 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 (MRM) [21, 10], steady state and/or transient analysis of the MRM produces the required measures of the original SRN.

3 SYSTEM DESCRIPTION

In cellular networks, a given geographical area is divided into a certain number of cells. In order to get the mutual interference beneath a tolerable threshold, each cell is allocated a fixed set of duplex channels which are different from those assigned to the adjacent cells. When a new call (NC) is attempted in a cell covered by a base station (BS), the NC is connected if an idle channel is available in the cell. Otherwise, the call is blocked. When a mobile station (MS) travels across the cell boundaries, the channel in the old serving cell is released, and an idle channel is required in the target cell, which would be the new serving cell. This process is called *handoff*. If an idle channel exists in the target cell, the handoff call (HC) continues nearly transparently to the user. Otherwise, the HC is dropped. In this investigation, we focus on both radio channel allocation, and radio channel failures. The RF system that we are investigating consists of a base site controller and many base stations.

The dropping of a handoff call (HC) is considered more severe than the blocking of a new call (NC). One method [5, 22] to reduce the dropping probability of HCs is to reserve a fixed number of channels exclusively for HCs. These exclusively reserved channels are referred as *guard channels*. For example, if the total number of RF channels is C and the number of the channels in the reserved channel pool is g, then the number of RF channels available for new calls is $C - g$.

The base site controller (BSC) handles radio channel allocation and many other functions. To date, the BSC only handles RF channel allocation for new calls and handoff calls, no RF channel failure of an ongoing call is recovered by BSCs. Whenever an RF channel fails, the call (FC) carried by the channel will be dropped. The RF channel failure can be categorized into permanent failure vs. transient failure and single channel failure vs. multiple channel failure. Permanent channel failure is caused by equipment impairment. Transient channel failure is caused by temporary conditions such as fading, blockage and etc. In this paper, only permanent single channel failures are considered. The models for transient and multiple channel failure and recovery are discussed in [23].

4 A PERFORMANCE MODEL FOR CHANNEL ALLOCATION

Before proposing our channel recovery scheme, we first present a pure performance model under the assumption that the channels in a wireless network never fail. Figure 1 shows an SRN of a performance model for channel allocation with guard channels.

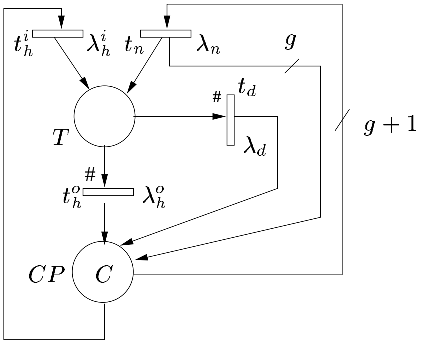

Figure 1: SRN of a performance model for channel allocation.

In Figure 1, place CP is the channel pool for the cell. Initially, there are C idle channels which are accessible for both the NCs and the HCs. Transitions t_n and t_h^i represent the arrivals of NCs and HCs respectively. Transition t_h^i is enabled with at least one idle channel in place CP. Otherwise, it is blocked. Transition t_n is disabled if there are less than $g+1$ channels in place CP. This is represented by the multiple input arc from place CP to transition t_n and the multiple output arc from transition t_n to place CP. The resulting effect is that when transition t_n fires, only one token is moved from place CP to place T. The number of tokens in place T is the number of channels currently being utilized in the cell. Transitions t_d and t_h^o respectively represent the departure of a call, either due to the termination of the call or due to the MS leaving the cell. The clearing rate for a *single* call is λ_d. The rate at which an MS leaves the cell is λ_h^o. Notice that transitions t_d and t_h^o have marking dependent firing rates. The *actual* firing rates for transitions t_d and t_h^o are $k\lambda_d$ and $k\lambda_h^o$ respectively, where k is the number of tokens in place T. The marking dependency is indicated by the # signs next to the transitions in Figure 1.

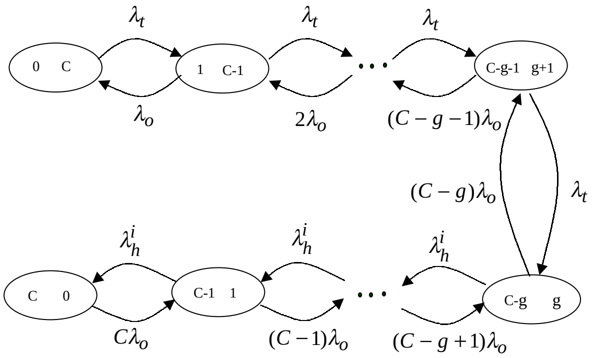

Figure 2: CTMC for the SRN model in Figure 1.

Let T_n denote the number of tokens in place T and consequently let $m = \{T_n, CP_n\}$ denote the marking of the SRN in Figure 1. The continuous time Markov chain (CTMC) for the SRN of the performance model is shown in Figure 2, where $\lambda_t = \lambda_n + \lambda_h^i$ and $\lambda_o = \lambda_d + \lambda_h^o$. With the underlying infinitesimal generator $\mathbf{Q}$ for the CTMC, numerical solution methods can be applied to get the desired different performance measures.

5 FIXED-POINT ITERATION

With different NC arrival rates, the corresponding HC arrival rates vary accordingly. To capture this dynamic behavior, we apply a fixed-point iteration scheme [11] to determine the HC arrival rates. The arrival rate λ_h^i of HCs should be equal to the actual throughput of transition t_h^o, denoted by Λ_{ho}. The value of Λ_{ho} can be calculated as follows,

$$\begin{aligned}\Lambda_{ho} &= \sum_{j\in\Omega}(\#[T_j])\lambda_h^o\pi_j(\lambda_h^i) \\ &= \sum_{j\in\Omega}(\#[T_j])\lambda_h^o\pi_j(\Lambda_{ho}). \qquad (1)\end{aligned}$$

By iteration, Equation (1) turns to be

$$\Lambda_{ho}^{s+1} = \sum_{j\in\Omega}(\#[T_j])\lambda_h^o\pi_j(\Lambda_{ho}^s). \qquad (2)$$

where Ω is the set of tangible markings of the SRN model, $\#[T_j]$ denotes the number of tokens in place T in marking (state) j, and $\boldsymbol{\pi}$ is the steady-state probability vector of the SRN model. Notice that $\boldsymbol{\pi}$ is a function of Λ_{ho}, because the firing rate of transition t_h^i should be equal to Λ_{ho}. When all the other parameters are fixed, an increase/decrease in λ_h^i (Λ_{ho}^s) implies an increase/decrease in Λ_{ho}^{s+1}. Therefore, Λ_{ho} is a monotone increasing/decreasing function of λ_h^i. In [11], Equation (1) is defined as the *fixed-point equation* and Λ_h^o is termed as the *iteration variable*. According to Theorem 2 in [11], a fixed point will exist if :

- The iteration function is a weighted sum of state probabilities and the weights are constants;
- The CTMCs underlying the SRNs are irreducible with more than one state.

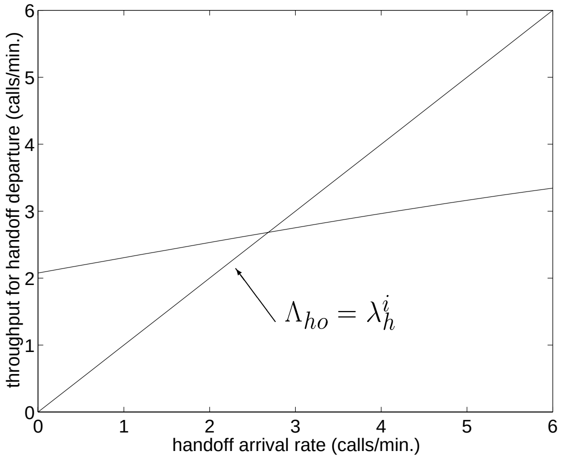

Figure 3: Handoff arrival rate (λ_h^i) vs. the throughput for handoff departure (Λ_{ho}).

It is easy to verify that the SRN model in Figure 1 satisfies the above two conditions. Therefore, a fixed-point exists for Equation (1). From Figure 3, we see that the fixed point solution is unique in our case. The numerical parameters for obtaining the figure are the same as those presented in Section 8.

6 TWO CHANNEL RECOVERY CASES

To increase RF channel availability, a channel recovery scheme is proposed in a progressive way. In Case I, to make the recovery model easier to understand, we assume that an idle channel never fails. In Case II, we release this assumption and present our recovery scheme in a comprehensive way. In our methods, a failed channel (FCh) is automatically switched by an idle channel, if one is available. Otherwise, the call with a failed channel is queued until an idle channel is available. Our methods are quite similar to the Automatic Protection Switching (APS) scheme, which is widely used to enhance the network integrity in the ATM networks [24]. In APS systems, a failed network component is switched to an identical spare component when the protection switch detects a failure. Since the spectrum is scarce, no spare channels are reserved exclusively for calls with failed channels. However, calls with failed channels are treated with the same priority as the HCs in the sense that both of them can access any available channel in the reserved channel pool.

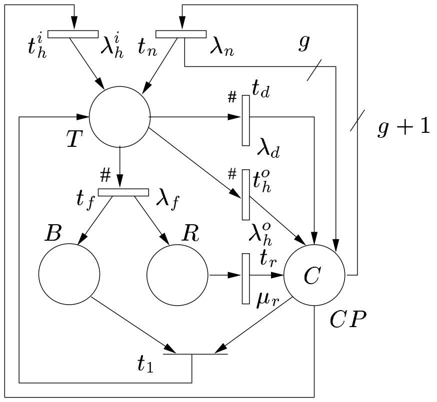

Figure 4: SRN for Case I.

6.1 A SIMPLIFIED RECOVERY SCHEME: CASE I

In this case, we assume that an idle channel is always in perfect condition for service. In other words, a channel can only fail when it is in service. A call fails when the channel it holds fails. We will release this assumption in Case II, the comprehensive recovery scheme.

In Figure 4, we show an SRN model for this scheme. Compared with the pure performance model in Figure 1, two places (B and R), two timed transitions (t_f and t_r) and one immediate transition (t_1) are added in Figure 4. Transition t_f represents the failure of a channel while it is in use. The failure rate for a *single* channel is λ_f. When a channel fails, the FC is switched to an idle channel if one is available. In this case, the FC is restored to service immediately. When the channel pool is empty, the FC is queued in the buffer B. As soon as an idle channel is available, an FC is restored instantly. The queued FCs are served by the first-in/first-out (FIFO) policy. The above process is represented by the immediate transition t_1 in Figure 4. In order to fire transition t_1, at least one token is required in both places CP (an idle channel) and B (an FC). In the mean time, the FChs are being recovered in place R under the FIFO policy with a single recovery facility. This is represented by transition t_r and the recovery rate is μ_r.

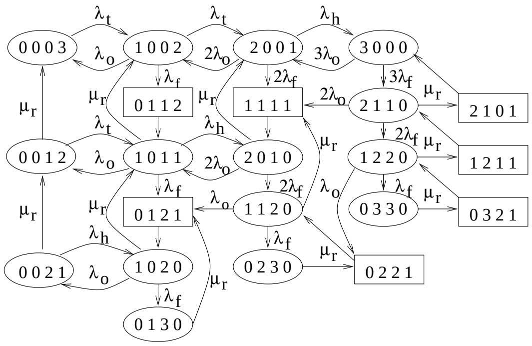

Figure 5: ERG for the SRN model in Figure 4.

Let T_n denote the number of tokens in place T and consequently let $m = \{T_n, B_n, R_n, CP_n\}$ denote the marking of the SRN in Figure 4. Then Figure 5 shows the $\mathcal{ERG}$ obtained from the initial marking shown in Figure 4, where $C = 3$ and $g = 1$. Vanishing markings are represented by rectangles and ovals are used to represent the tangible markings. The corresponding CTMC is shown in Figure 6.

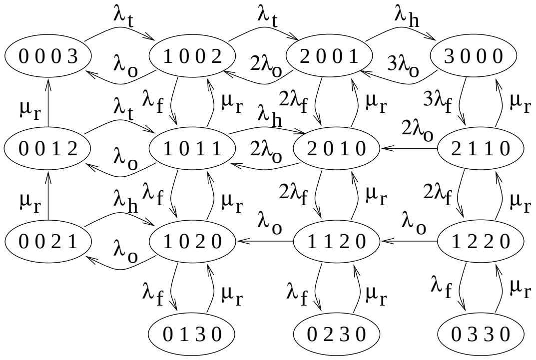

Figure 6: CTMC for the SRN in Figure 4.

Figures 5 and 6 show the potential problems we might encounter when the number of channels in place CP is large. In a practical environment, generally more than 20 channels are assigned to a BS. This would make the manual generation of the $\mathcal{ERG}$ and the associated CTMC very cumbersome, if not impossible. Thus a tool for automatically generating the $\mathcal{ERG}$ and solving the corresponding CTMC is needed. In this paper, we use the software package SPNP [20, 13] to specify and solve the SRN models.

6.2 A COMPREHENSIVE RECOVERY SCHEME: CASE II

In this case, we release the assumption that an idle channel never fails. We will consider the failure scenario for an idle channel. The comprehensive recovery algorithm (Figure 7) is composed of three major steps:

1. NC Req.: When a new call set-up process starts, a NC channel request is initiated for an RF channel from the available channel pool.
    - If there are more than g RF channels available in the pool, then an RF channel is obtained. Furthermore, if the RF channel is in working condition, then the call is set-up, and the number of available channels is decreased by 1. Otherwise, if the RF channel is bad, then a new RF channel request is issued and at the same time an RF channel repair request is issued. The new RF channel request may be put into the NC channel request queue if there is more than one NC RF channel request pending.
    - For an initial NC request, if there are less than g+1 idle RF channels, the call will be blocked.
2. FC/HC Req.: When an RF channel of an ongoing call fails or when a handoff request is initiated, an idle RF channel request is initiated from the channel pool.
    - If an RF channel is available, the RF channel is obtained. Furthermore, if the RF channel is in working condition, the number of available channels in the channel pool is decreased by 1, and the call continues as normal. Otherwise, if the RF channel is bad, a new RF channel request is issued and at the same time an RF channel repair request is issued. The channel request may be put into the FC/HC channel request queue if there is no channel available immediately.
    - For an initial FC/HC channel request, if there is no available RF channel, the call is dropped.
3. RF Channel Repair Update: Whenever an "RF channel repair completed" message is received, the RF channel allocation process increases the total number of available channels in the channel pool by 1.

An SRN model for the comprehensive recovery algorithm is developed as shown in Figure 8. On the right hand side of Figure 8, at place T_n^1, a NC has already obtained an idle channel, which is about to be tested for its working status. After testing, which is denoted by transition t_n^t, there are two possibilities.

1. With probability c_1 and the immediate transition t_6, the channel obtained is in working condition, and the NC is set-up in place T.

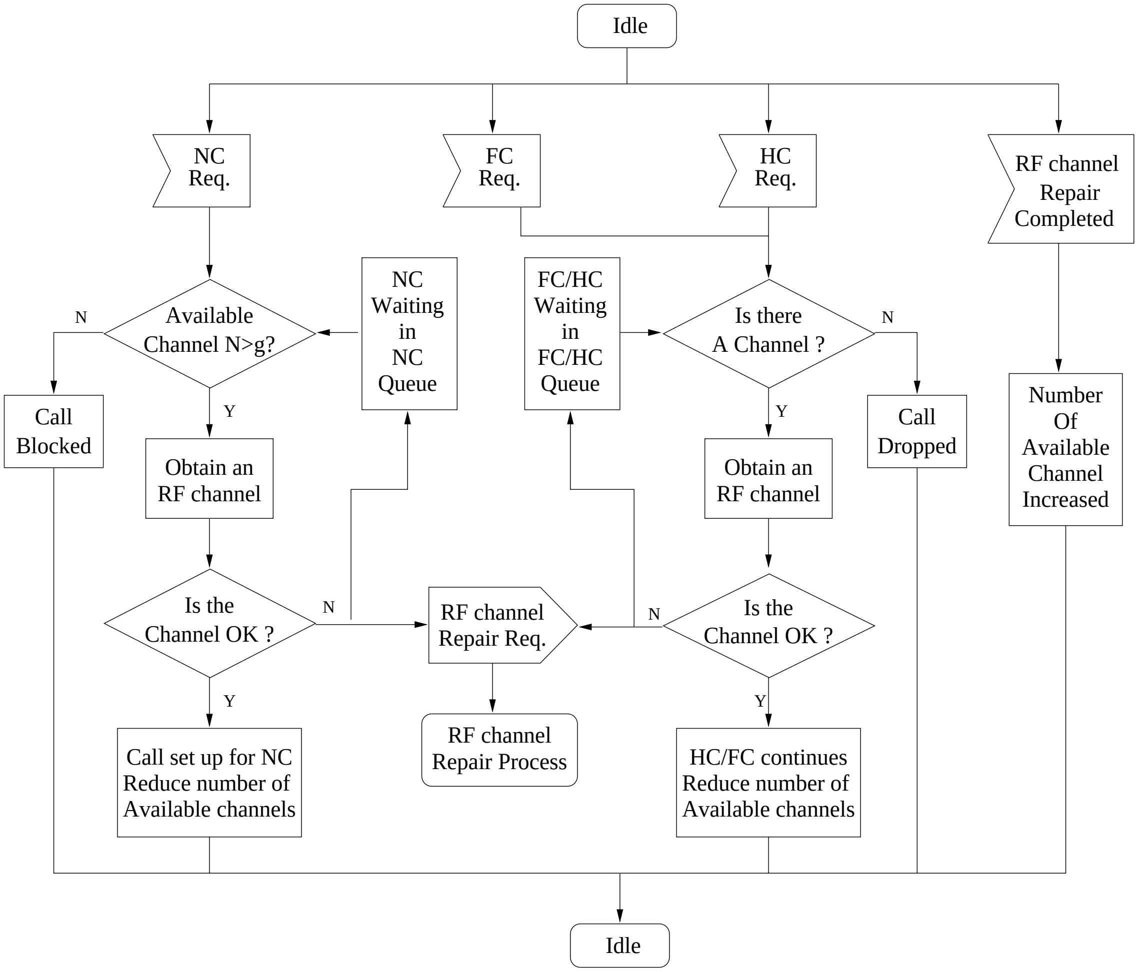

Figure 7: Flow chart for the RF channel allocation process.

2. With probability $c_2 = 1 - c_1$ and the immediate transition t_8, the channel obtained is in non-working condition, and the bad channel is deposited into the channel repair pool R. The NC is deposited into the waiting queue B_n. If there are more than g idle channels in the channel pool, no waiting is needed for the NC. The immediate transition t_4 is fired instantly. Again the idle channel is tested in place T_n^1. If there are less than $g+1$ channels in the channel pool, the NC needs to wait in the queue B_n.

Place T_n^2 acts as a temporary place. A token (representing a call/channel) coming into this place will go to either place T or places R and B_n immediately. On the left hand side of Figure 8, at place T_{hf}^1, an HC/FC has already grabbed an idle channel. Transitions t_{hf}^t, t_5 and t_7 have similar functions as their counterparts t_n^t, t_6 and t_8, respectively. If the idle channel obtained is in non-working condition, the HC/FC stays in place B before it fetches another idle channel through transition t_2. Another difference between Cases I and II is that an FC is dropped immediately (through the immediate transition t_3) if there is no idle channel in the channel pool. This is represented by the inhibitor arc from place CP to the immediate transition t_3. In this way, an FC is modeled with the same priority as the HC. Like place T_n^2, places T_{hf}^2 and F are temporary places. All the other places and transitions in Figure 8 have the same interpretation as in Figure 4.

6.3 A SIMPLIFIED SRN MODEL FOR CASE II

To reduce the number of states in the CTMC associated with the SRN in Figure 8, we simplify the SRN by removing the timed transitions t_{hf}^t and t_n^t and the corresponding places T_{hf}^2 and T_n^2. This is based on the fact that the RF channel checking time is very short. The simplified model is shown in Figure 9.

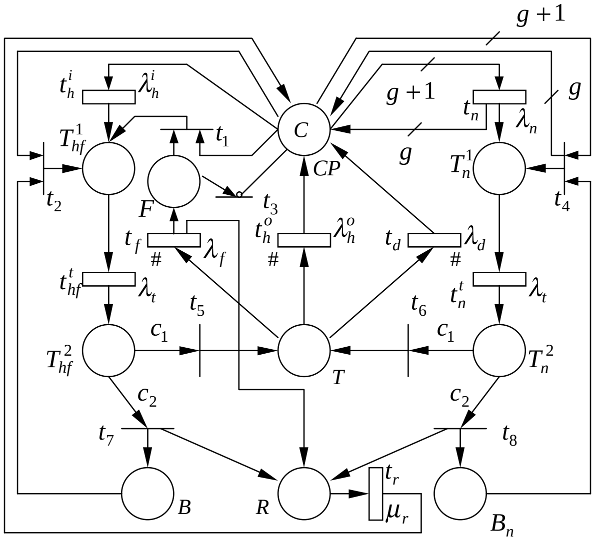

Figure 8: SRN for Case II.

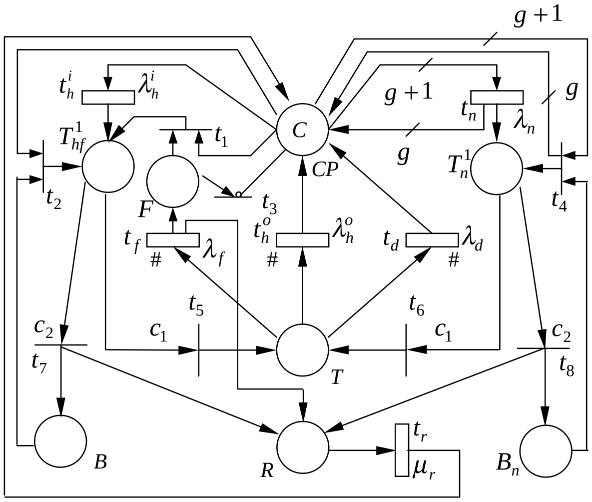

Figure 9: A simplified model of Figure 8.

6.4 AN SRN MODEL FOR RF CHANNEL ALLOCATION WITHOUT RECOVERY

For comparison purposes, an SRN model of RF channel allocation without recovery is developed and shown in Figure 10. The notations of places and transitions are respectively the same as those in Figure 9.

In the non-recovery method, once a channel carrying a call fails, no recovery is attempted. Thus place F and the related immediate transitions t_1 and t_3 as appeared in Figure 9 are not depicted in Figure 10. Furthermore, in the non-recovery scheme, if the idle channel that an NC/HC grabs is not functional, there is no retry. Consequently, places B and B_n, and the related immediate transitions t_2 and t_4 as depicted in Figure 9 are removed in Figure 10.

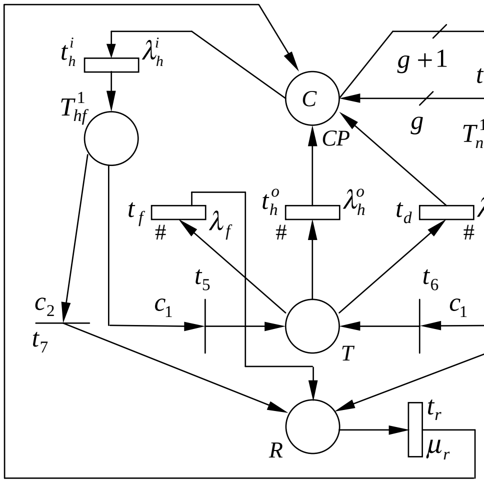

Figure 10: An SRN model for RF channel allocation without recovery.

7 MEASURES OF INTEREST

1. **Call Dropping Probability**: We denote the dropping probabilities for the recovery method and non-recovery methods as P_{dr} and P_{dn}, respectively. To calculate P_{dr}, the reward rate assignment is:

$$r_{dr}^{j} = \begin{cases} 1 & \text{if } [\#(CP_j)] = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Here r_{dr}^{j} is the reward rate for state j in the CTMC of SRN, and $\#(CP_j)$ represents the number of channels in place CP in marking (state) j. Thus a reward rate of 1 is assigned to the states where the channel pool is empty, and a reward rate of 0 is assigned to the other states. Then P_{dr} is calculated by

$$P_{dr} = \sum_{j \in \Omega} r_{dr}^{j} \pi_j,$$

where Ω is the set of all tangible markings and π_j is the steady-state probability of marking j. To calculate different measures of the system, we need to have different assignments of the reward rates.

The dropping probability for the non-recovery method (P_{dn}) is calculated through the following expression:

$$P_{dn} = P_{ch}^{e} + P_{hf}. \tag{4}$$

The first term of (4) is the probability that the channel pool is empty. The reward rates for P_{ch}^{e} are the same as in (3). The following formula is used to calculate P_{hf},

$$P_{hf} = \frac{\Lambda_{hf}}{\Lambda_t} P_R, \tag{5}$$

where Λ_{hf} is the *dropping* incoming rate for place R. It is contributed by the following two factors:

(a) A handoff call obtains a non-working idle channel, so it is dropped;

(b) The call is dropped while it is in talking status because the channel it holds fails.

In summary, Λ_{hf} is calculated through $\Lambda_{hf} = c_2 \cdot \Lambda_h^i + \Lambda_f$, where Λ_h^i and Λ_f are the throughputs of transitions t_h^i and t_f, respectively. The denominator Λ_t in (5) is the total incoming rate for place R. It is calculated through $\Lambda_t = c_2 \cdot \Lambda_h^i + c_2 \cdot \Lambda_n + \Lambda_f$, where Λ_n is the throughput of transition t_n. In (5), P_R is the sum of weighted state occupancy probabilities when the number of channels in place R is non-zero. The reward rate assignment for P_R is

$$r_R^j = \begin{cases} \#(R_j) & \text{if } [\#(R_j)] > 0, \\ 0 & \text{otherwise.} \end{cases}$$

2. **Call Blocking Probability**: We denote the blocking probabilities for the recovery and non-recovery methods as P_{br} and P_{bn}, respectively. To calculate P_{br}, a reward rate of 1 should be assigned to the states where the channel pool has less than $g+1$ channels, that is

$$r_{br}^j = \begin{cases} 1 & \text{if } [\#(CP_j)] \leq g, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The blocking probability for the non-recovery method (P_{bn}) is calculated through the following expression:

$$P_{bn} = P_{ch}^g + P_b. \tag{7}$$

The first term of (7) is the probability that the channel pool has less than $g+1$ channels. It has the same reward rate as in (6). The second term (P_b) in (7) represents the fact that a new call grabs a non-functioning channel and is blocked consequently. To calculate P_b, we use the following formula:

$$P_b = \frac{c_2 \cdot \Lambda_n}{\Lambda_t} P_R.$$

3. **Call Waiting Time**: After the initial access is granted, due to the channel failures, a NC or an ongoing call may need to stay in the queue B_n or B before it is assigned a working channel. We denote the expected waiting time for a NC in B_n as W_n. The expected waiting time in B for an ongoing call is represented by W_o. We assume that the probability of grabbing two faulty channels consecutively is zero. In other words, for a call waiting in the queue B_n or B, once it obtains a channel, it is assumed that the channel is in working status. In the following, we apply Little's formula [13] to calculate the values for W_n and W_o:

$$\begin{aligned} W_n &= \frac{E[\#(B_n)]}{c_2 \cdot \Lambda_n}, \\ W_o &= \frac{E[\#(B)]}{c_2 \cdot (\Lambda_h^i + (1 - P_t^3) \cdot \Lambda_f)}, \end{aligned}$$

where $E[\#(B_n)]$, $E[\#(B)]$ represents the average number of tokens in place B_n or B, respectively; Λ_h^i, Λ_f and Λ_n are the throughput of transitions t_h^i, t_f and t_n, respectively; P_t^3 is used to denote the probability that the immediate transition t_3 is enabled.

8 NUMERICAL RESULTS

Under the condition that all the neighboring cells are statistically identical and behave independently, the characteristics of the overall system can be captured by focusing on a single cell. In our system, all the cells are treated equivalently so that we can concentrate our attention on the performability aspects of the channel recovery schemes. By varying the traffic parameters, the presented schemes can also be applied to the cases where cells are located at the border of the covered area.

The grade of service (GOS) [25] is a measure of congestion, which is generally given as the probability of a call being blocked (for Erlang B) or as the probability of a call experiencing a delay greater than a certain queueing time (for Erlang C). Based on the blocking probabilities, we define the following:

- Light traffic: $GOS < 0.3\%$,
- Normal traffic: $0.3\% \leq GOS \leq 2\%$,
- Heavy traffic: $2\% < GOS$.

For the purpose of discussion, we assume that a set of $C = 26$ channels are assigned to each cell. The average travel time to cross from one cell to another ($1/\lambda_h^o$) is six minutes. The expected call holding time ($1/\lambda_d$) is 1.2 minutes. The mean time between failure ($1/\lambda_f$) for each channel is $80,000$ hours. The expected repair time ($1/\mu_r$) is 30 minutes. The failure probability (c_2) for an idle channel is approximated by $\lambda_f/(\lambda_f + \mu_r)$.

Now, we consider the problem of finding the optimal number (g) for guard channels such that a linear objective function of dropping and blocking probabilities is minimized. By using the guard channel policy, the dropping probability can be significantly reduced. However, reserving guard channels exclusively for HC/FC could result in blocking probability increase. To find a balance between these two GoS measures, we use the following function to consider the composite effect of dropping and blocking probabilities:

$$P_{bd}(g) = P_b(g) + w \cdot P_d(g). \tag{8}$$

Equation (8) is a weighted function of dropping and blocking probabilities, where $w > 1$, because we would like to give HC/FC a higher priority than NC. Our goal is to find a value of g such that P_{bd} is minimized. Numerical experiments are conducted to obtain the optimal value for g. According to the results shown in Figure 11, we choose to use $g = 1$ to obtain the numerical results in this section.

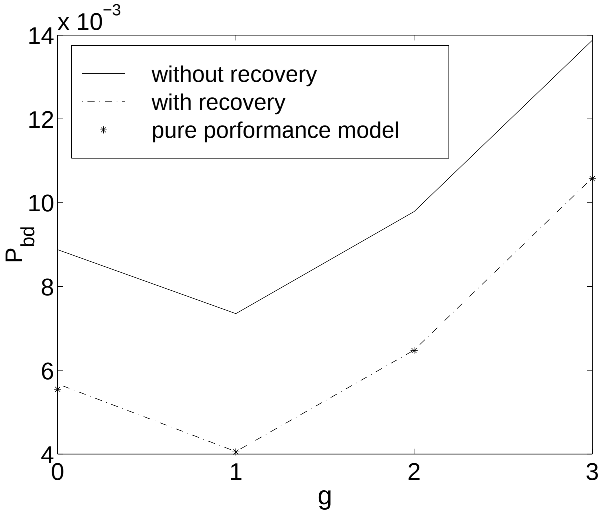

Figure 11: Number of guard channels (g) vs. (P_{bd}) when $w = 2$.

Table 1 shows the state space for the SRN models presented in this paper. From Figures 12, 13, 15, 16, 18 and 19, we notice that the numerical results for the recovery model and the performance model are nearly the same. Therefore, we conclude that the recovery method can nearly eliminate the dropped/blocked calls caused by channel impairment. In the following, we compare the performance results between the recovery and the non-recovery methods under different traffic loads.

Table 1: State space for the SRN models with $C = 26$ and $g = 1$.

Models	No. of Tangible Markings	No. of Nonzero Transitions
Performance Model	27	52
Comprehensive Recovery Model	655,200	4,125,797
Simplified Comprehensive Recovery Model	483	8337
Non-Recovery Model	378	1755

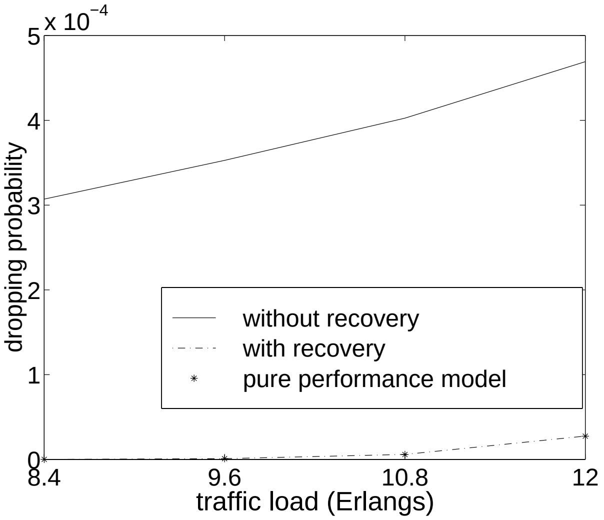

Figure 12: Dropping probabilities under light traffic loads.

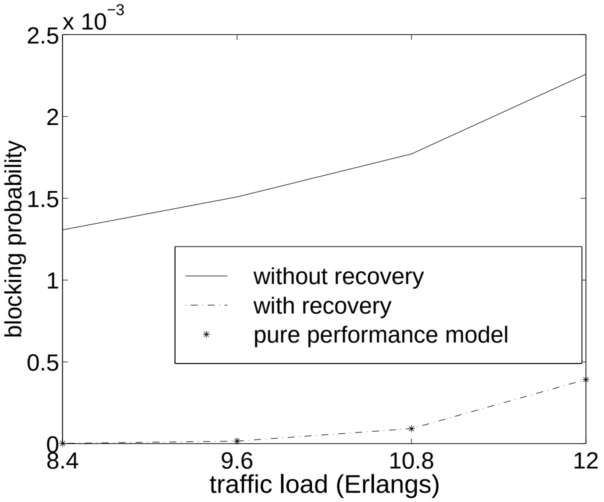

Figure 13: Blocking probabilities under light traffic loads.

- **Light Traffic:** When the traffic load ranges from 8.4 to 12 Erlangs (the corresponding NC arrival rate (NAR) ≤ 10 calls/min), the wireless network is undergoing light traffic (Figures 12 and 13). With the recovery method, the dropping probability is reduced to nearly zero ($\leq 2.758509 \times 10^{-5}$). When the traffic load is less than 10.8 Erlangs, the blocking probability can be reduced to nearly zero ($\leq 9.218801 \times 10^{-5}$). When the traffic load equals to 12 Erlangs, the blocking probability is reduced by 82.62%. The main reason for the nearly zero dropping/ blocking probabilities is that each of handoff calls, RF channel-failure calls, or new calls has multiple tries to obtain a working RF channel in the channel recovery method. For the original channel allocation algorithm without retries, any channel failure or temporary empty channel pool may cause dropped calls and/or blocked calls. Figure 14 shows the expected waiting times for the new calls and ongoing calls if their first tries are faulty channels. Under light traffic, $W_n < 2.2\,ms$ and $W_o < 0.9\,ms$. These waiting time would be negligible from the end-users' perspective.

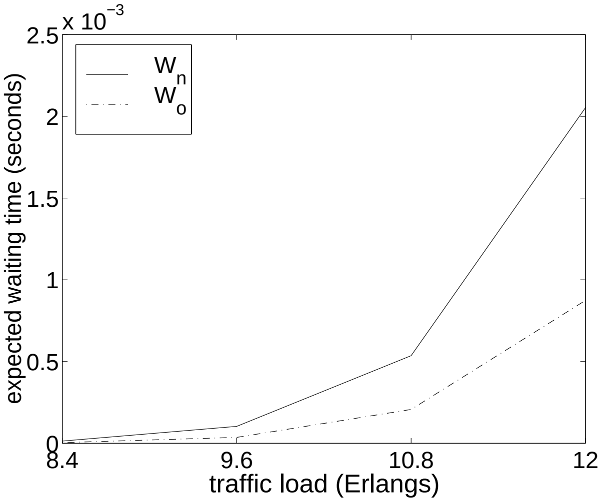

Figure 14: The expected call waiting times for the recovery method under light traffic loads.

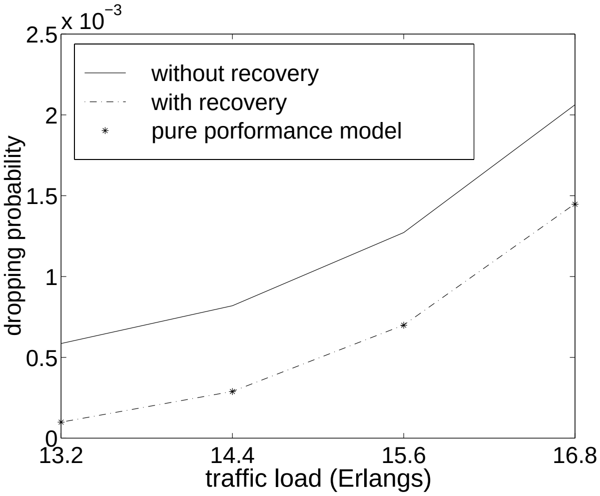

Figure 15: Dropping probabilities under normal traffic loads.

- **Normal Traffic:** When the traffic load is between 13.2 to 16.8 Erlangs (the corresponding NAR ≤ 14 calls/min), the traffic load is normal. Figures 15 and 16 show the dropping/blocking probabilities under normal traffic loads. The recovery method reduces the dropping and blocking probabilities at an average of 55.58% and 34.56%, respectively. The main reason is that each of the handoff calls or RF channel-failure calls has multiple tries to obtain a working channel in the recovery scheme. However, there is no second chance for the corresponding ones in the original non-recovery allocation method. Figure 17 shows the expected call waiting time under normal traffic, where the maximum channel recovery time is less than $60\,ms$ for new calls and less than $35\,ms$ for ongoing calls. These waiting times would be transparent for the end-subscribers.

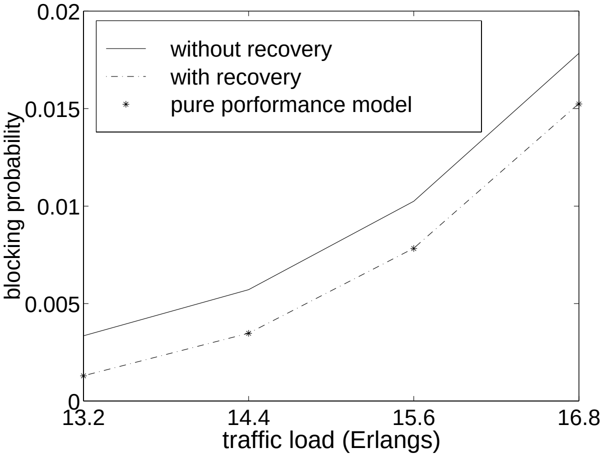

Figure 16: Blocking probabilities under normal traffic loads.

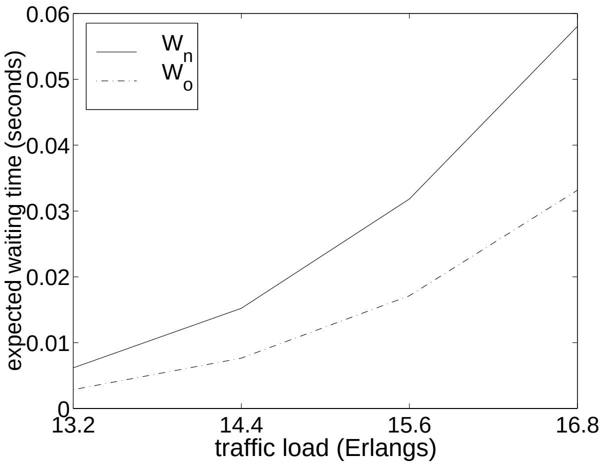

Figure 17: The expected call waiting times for the recovery method under normal traffic loads.

- **Heavy Traffic:** When the traffic load is between 18 to 21.6 Erlangs (the corresponding NAR ≤ 18 calls/min), the system is undergoing heavy traffic (Figures 18 and 19). The recovery method reduces the dropping and blocking probabilities at an average of 13.05% and 6.28%, respectively. We need to notice that as the traffic load increases, the chance of temporary empty channel pool increases accordingly. The dominant reason causing the blocked/dropped calls is now the heavy traffic, rather than the channel impairment. Therefore, under heavy traffic loads, the improvement of the channel recovery method can be limited. In complementary with the light and normal traffic, the expected call waiting times are shown in Figure 20. Although the waiting times are still tolerable, increasing the total number of RF channels should definitely be recommended to operating companies for larger revenue, more profit, and better customer satisfaction.

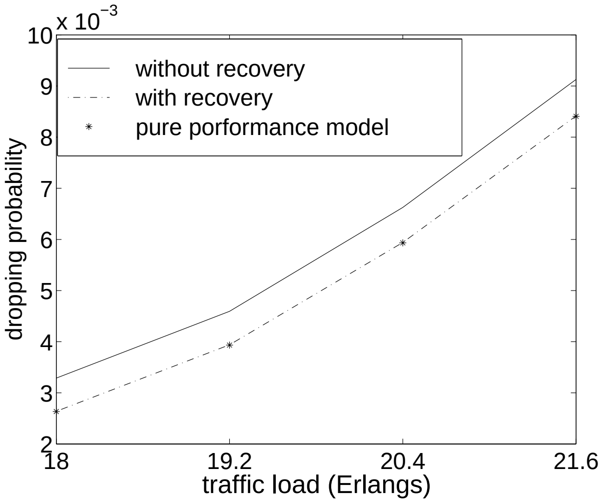

Figure 18: Dropping probabilities under heavy traffic loads.

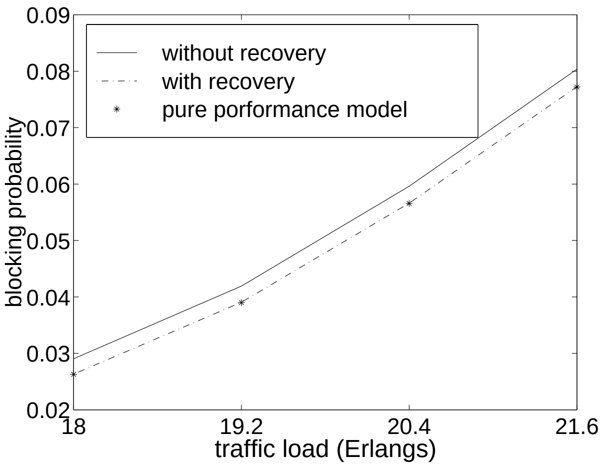

Figure 19: Blocking probabilities under heavy traffic loads.

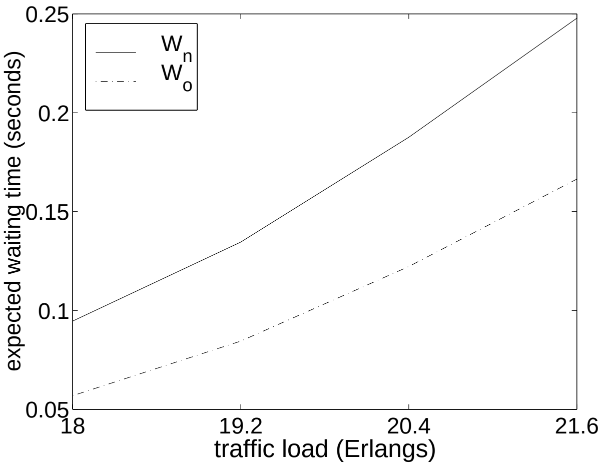

Figure 20: The expected call waiting times for the recovery method under heavy traffic loads.

9 CONCLUSION

In this paper, the SRN model for RF channel recovery is developed in a progressive way. First, a simple case is modeled under the assumption that an idle channel will not fail. Then, a comprehensive recovery model is presented by taking the idle channel failure scenario into account. To reduce the underlying state space for this comprehensive model, we simplified the model by combining/removing some transitions and places. For comparison purposes, an SRN model is also developed for the channel allocation method without recovery. To reflect the traffic pattern in a realistic way, a fixed-point iteration scheme is applied to capture the dynamic behavior of the handoff arrivals. The software package SPNP [20, 13] is applied to obtain the numerical results presented in this paper.

The results show that the comprehensive recovery scheme reduces the dropped calls and the blocked calls significantly under both light and normal traffics. Since the numerical results for the recovery model and the pure performance model are nearly the same, we conclude that the recovery model can eliminate the dropped/blocked calls caused by channel impairments. Under heavy traffic, although the recovery method can still reduce the dropping and blocking probabilities with limited percentage, increasing the total number of RF channels should be definitely recommended to the operating companies for better customer satisfaction. Because under heavy traffic, it is traffic rather than channel impairments that is the main contributor for the dropped/blocked calls. This paper addresses only permanent single channel failure and recovery. The models for transient and multiple channel failure and recovery are discussed in [23].

ACKNOWLEDGMENT

The authors would like to thank Professor F. Davoli (University of Genoa) and the anonymous reviewers for their helpful suggestions.

Manuscript received on . . .

REFERENCES

[1] G. Calhoun. *Wireless Access and the Local Telephone Network*. Artech House, Boston, MA, 1992.

[2] R. L. Freeman. *Radio System Design for Telecommunications*. John Wiley & Sons, New York. NY, second edition, 1997.

[3] P. Agrawal, D. K. Anvekar, and B. Narendran. Channel management policies for handover in cellular networks. *Bell Labs Tech. J.*, pages 97–110, 1996.

[4] L. J. Cimini, G. J. Foschini, C.-L. I, and Z. Miljanic. Call blocking performance of distributed algorithms for dynamic channel allocation in microcells. *IEEE Trans. Commun.*, 42(8):2600–2607, 1994.

[5] I. Katzela and M. Naghshineh. Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey. *IEEE Personal Communications*, 3(3):10–31, June 1996.

[6] Y.-B. Lin, S. Mohan, and A. Noerpel. Queueing priority channel assignment strategies for pcs hand-off and initial access. *IEEE Trans. Veh. Technol.*, 43(3):704–712, Aug. 1994.

[7] E. Del Re, R. Fantacci, and G. Giambene. Handover and dynamic channel allocation techniques in mobile cellular networks. *IEEE Trans. Veh. Technol.*, 44(2):229–237, May 1995.

[8] C. H. Yoon and C. K. Un. Performance of personal portable radio telephone systems with and without guard channels. *IEEE J. Select. Areas Commun.*, 11(6):911–917, Aug. 1993.

[9] J. F. Meyer. Performability: a retrospective and some pointers to the future. *Performance Evaluation*, 14(3-4):157–196, Feb. 1992.

[10] K. S. Trivedi, J. K. Muppala, S. P. Woolet, and B. R. Haverkort. Composite performance and dependability analysis. *Performance Evaluation*, 14(3-4):197–215, Feb. 1992.

[11] V. Mainkar and K. S. Trivedi. Sufficient conditions for existence of a fixed point stochastic reward net-based iterative models. *IEEE Trans. Software Engineering*, 22(9):640–653, Sep. 1996.

[12] M. Ajmone-Marsan, D. Kartson, G. Conte, and S. Donatelli. *Modelling with Generalized Stochastic Petri Nets*. John Wiley & Sons, Inc., New York, NY, 1995.

[13] G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi. *Queueing Networks and Markov Chains, Modeling and Performance Evaluation with Computer Science Application*. John Wiley & Sons, New York, NY, 1998.

[14] S. Hunter, T. Phillip, and K. S. Trivedi. Combined performance and availability analysis of a switched network application. In *Proceedings of the International Conference on Communications (ICC'97)*, Montréal, Québec, Canada, June 8-12 1997.

[15] O. C. Ibe, H. Choi, and K. S. Trivedi. Performance evaluation of client-server systems. *IEEE Transactions on Parallel and Distributed Systems*, 4(11):1217–1229, November 1993.

[16] F. J. Jaimes-Romero, D. Muñoz-Rodríguez, C. Molina, and H. Tawfik. Modeling resource management in cellular systems using Petri nets. *IEEE Trans. Veh. Technol.*, 46(22):298–312, May 1997.

[17] Y. Ma, J. J. Han, and K. S. Trivedi. Call admission control for reducing dropped calls in code division multiple access (cdma) cellular systems. In *Proceedings of IEEE INFOCOM 2000*, Tel-Aviv, Israel, March 2000.

[18] C. Molina, N. Jain, and K. Basu. Performance model of cellular data on American systems. In *Proceedings of 1996 IEEE 46th Vehicular Technology Conference (VTC'96)*, Atlanta, GA, 28 April - 1 May, 1996.

[19] C.-Y. Wang, D. Logothetis, and K. S. Trivedi. Transient behavior of atm networks under overloads. In *Proceedings of the IEEE INFOCOM 96*, pages 978–985, San Francisco, CA, USA, March 1996.

[20] G. Ciardo, J. K. Muppala, and K. S. Trivedi. SPNP Users Manual, Ver. 5.01. Technical report, Duke University, Durham, NC, 1998.

[21] G. Ciardo, A. Blakemore, Jr. P. F. Chimento, J. K. Muppala, and K. S. Trivedi. Automated generation and analysis of Markov reward models using stochastic reward nets. In C. Meyer and R. Plemmons, editors, *Linear Algebra, Markov Chains and Queuing Models*, volume 48, pages 145–191. Springer-Verlag, 1993.

[22] N. D. Tripathi, J. H. Reed, and H. F. Vanlandingham. Handoff in cellular systems. *IEEE Personal Communications*, 5(6):26–37, Dec. 1998.

[23] Y. Ma, J. J. Han, and K. S. Trivedi. A channel recovery method in TDMA wireless systems. In *Proceedings of the 50th IEEE International Vehicular Technology Conference (VTC'99–Fall)*, pages 1750–1754, Amsterdam, The Netherlands, Sep. 1999.

[24] T.-H. Wu and N. Yoshikai. *ATM Transport and Network Integrity*. Academic Press, San Diego, CA, 1997.

[25] T. S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, Upper Saddle River, NJ, 1996.