Performability Modeling and Analysis for Wireless Cellular Networks with Prioritized Handoff Scheme

Abstract:

Nowadays, both of network operators and regulators are concerning about adding new end user's features. These features may be added to future networks and current deployed ones. In order to achieve that, network performance as well as its availability must be taken into consideration side-by-side for having a successful network operation. The process of coupling of them will be beneficial in the characterization process of wireless communication systems. This kind of wireless network assessment is defined as performability analysis. In this paper, a wireless network is modeled with break-downs and multi-repair techniques. Unlike the previous studies, for such systems, an exact model and solution approach is used for the performability assessment, and handoff priority issues are considered together with multi-recovery channel availability issues. Numerical results have been obtained and presented for various performability parameters, such as mean queue length (MQL), new call blocking probability (Pb), and handoff forced termination probability (Pfh). The presented model investigates the deployment of handoff priority scheme, in conjunction with the different arrival rates for different system capacities. The obtained results are presented comparatively to validate the obtained system behavior.

Index Terms—Availability
Performability Analysis and Modeling in Cellular Networks with Handoff Fractional Guard Channel Policy

Hesham ElBadawy, Member, IEEE

Abstract—Nowadays, wireless networks have become paramount importance to handle the categories of customer’s calls. Specially concern must be taken into consideration for the handoff category over the category of fresh calls. The impact of handling such networks must tackle not only teletraffic analysis, but also, the case of wireless channel unavailability. This kind of wireless network assessment is defined as performability analysis. In this paper, a wireless network is modeled with break-downs and multi-repair techniques. A queueing model will be used in conjunction with the deployment of Fractional Guard Channel (FGC) call admission policy. In addition, the current paper provides a complete model description and the detailed mathematical derivation and its numerical solution for this kind of untractable problems. Numerical results have been obtained and presented for various performability parameters, such as mean queue length (MQL), new call blocking probability (Pb), and handoff forced termination probability (Pfn). The presented model investigates the deployment of handoff priority scheme, in conjunction with the different arrival rates for different system capacities. The obtained results are presented comparatively to validate the obtained system behavior.

Index Terms—Wireless Performability; Availability of cellular network; Handoff priority scheme; Performability modeling; Fractional Guard Channel.

I. INTRODUCTION

Cellular systems divide the service area into multiple adjacent cells where each cell has its own base station (BTS). Mobile subscribers (MSs) communicate via air interface with BTSs, and mobility is one of the major issues in the performance characterization of these wireless communication systems. A steady-state teletraffic analysis has been provided via many analytical models. In addition, performance of such networks with the deployment of handoff priority schemes has been presented in [1]-[5]. In [1]-[8], the wireless networks performance has been tackled mainly from the pure performance point of view. The networks were assumed to have no failures or service unavailability situations. As a matter of fact, wireless communication systems encounter failures due to various reasons (such as failure of software, hardware, environmental conditions, human error, or any combination of these factors [14]). Queuing models are extensively used in the modeling process of different communication networks. Queuing scenarios where the service rates depend on the amount of present traffic are presented in [8],[9]. In [7]-[16], the guard channel policy [3] is the well-known technique to cope with the problem by giving a priority for handoff calls over fresh calls. Fractional Guard Channel (FGC) scheme is modeled for both of new (or fresh) calls as well as the handoff calls with the different level of priorities. The performability model was derived without considering handoff prioritization scheme, and single repair methodology was assumed. Differently from the previously published work in [3],[4],[7],[16], the current paper aims to investigate and model the performability of a wireless network with handoff priority scheme as well as the multi breakdowns and repair mechanism. However the failure/repair phenomenon of calls in wireless networks with FGC was not discussed in the literature up until now. The consideration of this phenomenon allows the investigation of important performance measures related to the quality of services (QoS) experienced by subscribers such as the average number of calls in the buffer, the probability of fresh and handoff calls leaving the system due to the control policy, and the probability that fresh and handoff calls are forced to the repairs’ buffer. Otherwise, the handoff calls will be accepted until the system is fully occupied with N channels. Handoff calls will be dropped due to system unavailability or no available resources. In addition, the current work analyzes a wireless network of single-channel queues with state-dependent inter-arrival times and service rate are taken into account. Systems which are prone to breakdown are investigated in conjunction with handoff priority schemes. In addition, service rates which are dependent on the number of calls in the system will be taken into consideration. Quasi birth death (QBD) process is employed together with the spectral expansion (SE) method to derive the exact solution of a two dimensional Markov model. The novelty of the current work is the exploration and the adoption of the SE methodology to analyze the performability of a wireless network with handoff priority scheme and multi breakdowns and repairs. The presented mathematical analysis provides a complete performability model, and its numerical solution. The assessment parameters such as: mean queue length (MQL), new call blocking probability (Pb), and handoff forced termination probability (Pfn) will be determined.

The rest of paper is organized as follows: in section II, model description is provided. In section III, two dimensional Markov process model and, the spectral expansion methodology are presented as an exact approach for system’s performability assessment. In section IV, the numerical results will be obtained and compared to the previously published work in...
In addition, the performability assessment parameters are illustrated and analyzed for different system capacities and operational scenarios such as: the percentage of guard channels and the new call arrival rate. Finally, conclusion and directions for future work are provided in section V.

II. SYSTEM DESCRIPTION

The current paper considers a particular cell in a huge steady state cellular mobile communications network with infinite user population. So, the model concentrates on only one cell, which has \( N \) channels, with a queuing capacity \( H_q \) (to avoid the unavailability situations). Thus, the maximum number of calls allowed into the system is equal to the number of calls assigned with the channels in case of a fully operational system plus the queuing capacity. The maximum system’s capacity (i.e., maximum number of calls in the system) is denoted by \( L \); where \( L = N + H_q \). A single bounded (finite) queue is served by \( N \) identical parallel channels (servers). Each channel goes through alternating periods of being operative and inoperative, independently of the others and of the number of calls in the system. The operative and inoperative periods are exponentially distributed with parameters \( \xi \) and \( \eta \), respectively [13]. Calls arrive according to Interrupted Poisson Process (IPP) controlled by a process denoted \( \lambda(t) \); where: \( \{I(t) | i \in I(t), t \geq 0, 0 \leq i \leq N\} \). Calls are taken for service from the head of the queue, one at a time, by available operative channels. The required service times are distributed exponentially with parameter \( \mu \). An operative channel cannot be idle if there are calls waiting to be served. A call whose service is interrupted by a channel breakdown is returned to the head of the queue. Similar to the previous studies which was presented in [3], [4], [6], [7], [12], and [14] the inter-arrival times of the incoming call requests and the service times of the calls served by channel \( i(0, 1, 2, \ldots, N) \) are assumed to follow exponential distributions. The operative periods and repair times of channel \( i \), are also distributed exponentially similar to the studies presented in [3], [4], [14], and [15]. Two different kinds of arrival rates are defined for fresh calls and handoff calls with mean arrival rates \( \lambda_f \) and \( \lambda_h \), respectively. The call arrivals can be assigned in any channel in the cell. Otherwise, in case of channel unavailability, the ongoing call (new or handoff) is added to the queue [7] until the handoff threshold \( (N - N_g) \) is not achieved. Otherwise, the handoff calls will be served or queued but the fresh calls will be blocked. Figure (1), illustrates the proposed queuing model for the performance analysis. The presented model is a modified one from that presented in [7]. Let \( \lambda_T \) be defined as the total arrival rate of calls in the cell, where:

\[
\lambda_T = \lambda_f + \lambda_h \tag{1}
\]

Hence, an explicit ergodicity condition is obtained by requiring offered load to be less than the average number of operative channels:

\[
\frac{\lambda_T}{\mu} < \frac{N \eta}{\xi + \eta} \tag{2}
\]

The time that an MS spends in the cell (dwell time) is assumed to be exponentially distributed with mean \( 1/\mu_{ed} \), where, \( \mu \), \( E_V \) is the average velocity of the MS, \( F \) is the perimeter of cell, and \( R \) is the cell radius [7]. An exponentially distributed call holding time, \( T_c \), is also assumed with mean \( 1/\mu \). Then, the total channel holding time, \( T \), can be defined as an exponentially distributed with mean \( E[T] = 1/\mu \) given that for a hexagonal cell shape with radius of \( R \), the expectation of the channel holding time is given by [19]:

\[
\mu = \mu_c + \mu_{ed} \tag{3}
\]

\[
T = \frac{\pi R \sqrt{3}}{4 E_V + \pi R \mu_c \sqrt{3}} \tag{4}
\]

In the current paper, a multi repair facilities are assumed for all channels. The channel failure and repair behavior are shown in figure (2), for a system with \( N \) channels.

From a pure system performance perspective, the state transition diagram of the considered wireless cellular systems is illustrated in figure(3). This is in consistence with the previously proposed models in [12], [11].

In the current work, since the system can be approximately modeled by a quasi birth death (QBD) process, some existing methods can be used such as the improved version [13] of the matrix-geometric method (MGM) [14], the spectral expansion method (SE) [15]. We will show that we can deploy a faster algorithm than the matrix-geometric method. Figure (3), shows both of arrival and service rates are dependent on the number of calls in the system.

III. SPECTRAL EXPANSION SOLUTION

The Fractional Guard channel [2] policy is defined as follows. When \( I(t) = i \) \( (0 \leq i \leq N) \), a fresh call or repaired call
and handoff calls are accepted with probability 1. The system is described by the two-dimensional continuous time Markov process \((I(t), J(t))\), where \(I(t)\) represents the number of repaired servers (channels in the context of this paper) and \(J(t)\) is the number of customers served in the system. The steady state probabilities are denoted by \(\pi_{i,j}\). Vector \(V_j\) is considered as follows: \(V_j = (\pi_{0,j}, \pi_{1,j}, \pi_{2,j}, \ldots, \pi_{N,j})\). Repaired queues have been used to model the queueing problem of fresh and handoff calls in cellular mobile networks [4]. The fact that the repair rate depends on the number of repaired calls waiting in the system leads to an analytically intractable model [5], [12]. Therefore, approximation procedures should be used to compute the performability of the system. The performability of the cellular network which has been described in section II, the current section provides complete and detailed analysis of the cellular network which has been described in section II.

II. Methodology and Its Adopted for the Current Markov Process

The proposed methodology and its adoption for the current Markov process is described by the two-dimensional continuous time Markov model. The second stage, via four different stages. The first stage, is to construct the QBD process, an irreducible Markov process a \((QBD process)\). The third one will present the spectral expansion methodology and its adoption for the current Markov process. The third one will consider the evaluation criteria to find out the state transition probability matrix. Finally, the fourth one will provide the performability assessment parameters.

A. Markov Process

Following the presented methodology by [2], [6], the present case will be investigated and modeled. The state of the system at time \(t\) can be described by a pair of integer valued random variables, \(I(t)\); where: \(\{I(t) \mid i \in I(t), t \geq 0, 0 \leq i \leq N\}\); specifying the channel availability. So that, it is assumed that there are \(N+1\) channel configurations (i.e. failures and repairs). These \(N+1\) configurations are the possible state variables of the model. On the other hand, \(J(t)\) is the other process that specifies the number of calls (or services) in the system including the one(s) in service; where \(\{J(t) \mid j \in J(t), t \geq 0, 0 \leq j \leq L\}\). Both of \(I(t)\) and \(J(t)\) are an irreducible Markov process. The proposed (lattice strip) will be presented by two dimensional \((2D)\) Markov process, which is consisting of the two perpendicular models. In the horizontal direction (the availability model), the possible state variables in the system are presented. Whereas, in the vertical direction (the performance model) is considered to represent the number of calls in the system. Then, an irreducible 2D Markov process a \((QBD process)\), \(Z = \{I(t), J(t); t \geq 0\}\) is constructed. The size of \(Z\) is \(\{N+1\} \times \{L+1\}\).

Let, \(A\) is the matrix of instantaneous transition rates in the availability model. On the other hand, let matrices \(B\) and \(C\) are transition matrices for one-step upward and one-step downward transitions, respectively in the performance model. The characteristic equation of the system is deduced from the general two dimensional Markov model with modulated queues by means of three types of transitions:

1) \(1^{st}\) type :: Phase transitions which leave the queue unchanged: from state \((i, j)\) to state \((k, j)\), where; \((0 \leq i; k \leq N; i \leq k)\), with instantaneous repair/failure rate \(a_i(i, k)\).

2) \(2^{nd}\) type :: Transitions associated with the arrival of a call, possibly together with a change of phase: from state \((i, j)\) to state \((k, j+1)\), where \((0 \leq i; k \leq N; j > 0)\), with instantaneous arrival rate \(b_j(i, k)\).

3) \(3^{rd}\) type :: Transitions associated with the departure of a call, possibly together with a change of phase: from state \((i, j)\) to state \((k, j-1)\), where \((0 \leq i; k \leq N; j > 0)\), with instantaneous departure rate \(c_j(i, k)\).

\[ A_j = \begin{bmatrix} 0 & N\eta & 0 & \cdots & 0 & 0 \\ \xi & 0 & (N-1)\eta & \cdots & 0 & 0 \\ 0 & 2\xi & 0 & \cdots & 0 & 0 \\ 0 & 0 & 3\xi & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 2\eta & 0 \\ 0 & 0 & 0 & (N-1)\xi & 0 & \eta \\ 0 & 0 & 0 & \cdots & N\xi & 0 \end{bmatrix}; 0 \leq j \leq L \]  

\[ B_j = \begin{bmatrix} \lambda_T & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_T & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_T & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_T & 0 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_T \\ 0 & 0 & 0 & \cdots & 0 & 0 & \lambda_T \end{bmatrix}; 0 < j \leq (N-N_0) \]  

\[ B_j = \begin{bmatrix} \lambda_h & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_h & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_h & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_h & 0 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_h \\ 0 & 0 & 0 & \cdots & 0 & 0 & \lambda_h \end{bmatrix}; (N-N_0) < j \leq L \]  

C matrix is dependent on the number of calls for \((j = 0, 1, \ldots, L)\). \(C_j\) has two defining matrices: first if \((0 < j \leq N)\), so that the number of calls in the system is less than the total number of available channels, a channel is assigned for each call. Therefore the downward transition rate is chosen as the minimum of number of calls and number of available
Let $D$ denote the number of channels.

$$C_j = \begin{bmatrix}
0 & \min(1,j)\mu & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & \min(N,j)\mu
\end{bmatrix}$$  \hspace{1cm} (8)

On the other hand, if $N < j \leq L$ then, the operative channels can provide service only for handoff calls. So, all of the available channels are assigned to the incoming handoff calls and the other calls in the queue with service rate as $\mu, d$ \cite{7}, \cite{15}. Let $A_j = [a_j(i,k)], B_j = [b_j(i,k)]$ and $C_j = [c_j(i,k)]$, respectively (the main diagonal of $A_j$ is zero by definition; also, $C_0 = 0$ by definition).

### B. Spectral Expansion Methodology

In order to solve this system of equations is done via two different methodologies, namely: Spectral Expansion (SE), and Matrix Geometric Method (MGM). The main operational difference between these two approaches is that spectral expansion involves the computation of eigenvalues and eigenvectors of a given matrix, whereas the matrix-geometric method requires the determination of an unknown matrix by solving a matrix equation. When comparing the spectral expansion and the matrix-geometric solutions, a strong case can be made for the former. Spectral expansion offers considerable advantages in efficiency, without undue penalties in terms of numerical stability. The speed gains are especially pronounced, and indeed appear to be unbounded, when the system approaches saturation. However, further comparisons, perhaps with different methods for computing the unknown matrix, or with other methods such as \cite{13}. To evaluate the impact of that difference, a series of experiments where performed in \cite{14}. Knowing that, the wireless networks have limited radio resources. So, SE will be more faster, and have sufficient stability. In addition, SE will not be proud to instability for huge matrix spaces, which are rarely found in wireless network models. In order that, the current work will deploy the SE for obtaining the performability metrics.

Now, the solution of the current model will be obtained via the balance equations and the state equilibrium probabilities. The left-hand side of (10) gives the total average number of transitions out of state $\pi(i,j)$ per unit time (due to changes of phase, arrivals and departures), while the right-hand side expresses the total average number of transitions into state $\pi(i,j)$. These balance equations can be rewritten more compactly by using vectors and matrices. By recalling the row vectors of probabilities corresponding to states with $j^{th}$

Let $A_j, B_j$ and $C_j$, respectively, then:

$$D_j^A(i,j) = \sum_{k=0}^{N} A_j(i,k)$$  \hspace{1cm} (12)

$$D_j^B(i,j) = \sum_{k=0}^{N} B_j(i,k)$$  \hspace{1cm} (13)

$$D_j^C(i,j) = \sum_{k=0}^{N} C_j(i,k)$$  \hspace{1cm} (14)

Then, (10) may be re-written as follows:

$$V_j \left[ D_j^A + D_j^B + D_j^C \right] = V_j - 1 B_{j-1} - V_j A_j + V_{j+1} C_{j+1}$$  \hspace{1cm} (15)

by definition, $V_{-1} = 0$ and $B_{-1} = 0$. Also, the boundary conditions are as follows:

$$V_0 \left[ D_0^A + D_0^B \right] = V_0 A_0 + V_1 C_1$$  \hspace{1cm} (16)

$$V_L \left[ D_L^A + D_L^C \right] = V_{L-1} B_L + V_L A_L$$  \hspace{1cm} (17)

In addition to the normalization property that all probabilities must sum up to 1:

$$\sum_{i=0}^{N} \sum_{j=0}^{L} \pi(i,j) = 1$$  \hspace{1cm} (18)

The first step of any solution method is to find the general solution of the set of balance equations with constant coefficients, \cite{12}. Rewrite (13) in the form of a homogeneous vector difference equation of order 2:

$$V_j Q_0 + V_{j+1} Q_1 + V_{j+2} Q_2 = 0$$  \hspace{1cm} (19)

where,

$$Q_0 = B; \quad Q_1 = A - D^A - D^B - D^C; \quad Q_2 = C$$  \hspace{1cm} (20)

Let $Q(x)$ and $Q(y)$ be the "characteristic matrix polynomial" and its "complementary polynomial associated" respectively. So, the difference equations may be represented as follows:

$$Q(x) = Q_2 x^2 + Q_1 x + Q_0$$  \hspace{1cm} (21)

$$Q(y) = Q_2 + Q_1 y + Q_0 y^2$$  \hspace{1cm} (22)

The solution for $Q(x)$ is denoted by $x_k$ as the 'generalized eigenvalues', whereas $u_k$ is denoted as the corresponding 'generalized left eigenvectors'. In other words, these are quantities which satisfy:

$$\| Q(x) \| = 0, \quad \text{and} \quad u_k Q(x_k) = 0; \quad k = 1, 2, \ldots, N$$  \hspace{1cm} (23)

In the present model, $Q(x)$ is singular. So, the complementary function of $Q(x)$ will be used, $Q(y)$ will be used instead of $Q(x)$. The solution for $Q(y)$ is denoted by $y_k$ as the 'generalized eigenvalues', and $\psi_k$ as corresponding 'generalized left eigenvectors'. In other words, these are quantities which satisfy:

$$\psi_k Q(y_k) = 0; \quad \| Q(y) \| = 0, \quad \text{and} \quad k = 0, 1, 2, \ldots, N$$  \hspace{1cm} (24)
Hence, the spectral solution of $V_j$ is generalized as follows:

$$V_j = \sum_{k=1}^{N+1} \beta_k \psi_k \psi_k^{L-j}; \ |y_k| > 1$$ (25)

Where $\beta_k$ is an arbitrary constant (may be complex). In order to compute the eigenvalues of $Q(y)$, $\psi_k$ outside the unit disk, and its corresponding left eigenvectors $\psi_k$. The scalar polynomial, $\|Q(y)\|$, is solved in order to find its roots, which may be very inefficient for large system capacities.

$$\psi_k \left[ Q_2 + Q_1 y + Q_0 y^2 \right] = 0$$ (26)

By using the double sized matrix $G$ to reduce the quadratic eigenvalue-eigenvector problem to a linear one of the form [3]:

$$y G = \psi \psi$$ (27)

N.B; $G$ is a matrix whose dimensions are twice as large as those of $Q_0$, $Q_1$ and $Q_2$. This linearization can be achieved, if the matrices in (20) are nonsingular. Indeed, $C$ matrix is singular one, so we may achieve the required linearization by means of multiplying both sides by $B^{-1} = Q_0^{-1}$, where:

$$Q_0^{-1} = B_j^{-1} = \frac{1}{\lambda_s} I; \quad \lambda_s = \begin{cases} \lambda_f, & 0 < j < (N - N_y) \\ \lambda_h, & (N - N_y) \leq j \leq L \end{cases}$$ (28)

Where, $I$ is the unity diagonal matrix. So, (26) may be rewritten as in the following form[14]:

$$\begin{bmatrix} \psi & y \psi \end{bmatrix} \begin{bmatrix} 0 & -H_o & I \\ I & -H_1 \end{bmatrix} = y \begin{bmatrix} \psi & y \psi \end{bmatrix}$$ (29)

where,

$$\begin{bmatrix} H_o \\ H_1 \end{bmatrix} = \left( A - D^A - D^B - D^C \right) B^{-1}$$ (30)

Hence, $\psi_k$ is obtained as the lowest eigenvalue just outside the unity disk, and $\psi_k$ is its associated eigenvector. The vectors ($V_{L-1}, V_{L-2}, V_{L-3}, \ldots, V_0$) can be expressed in terms of $V_L$, by using the recursion of equations(15) through (18).

C. State Transition Matrix

After the linearization of the eigenvalue problem, the implicit restarted Arnoldi method may be deployed. Then, the coefficients $\beta_0, \beta_1, \beta_2, \cdots, \beta_N$ are computed as the solution of linear equation of the $V_j$. So, the current work may be computed for different network sizes and loads. So, following the approximation algorithm , which had been proposed in [16]. So, for the operational with light and moderate loads, the steady state probabilities can be expressed as follows:

$$V_j = \beta_{L-1} \psi_{L-1} y_{L-1}^{L-j} + \beta_L \psi_L y_{L}^{L-j}$$ (31)

$$V_0 = \sum_{k=0}^{N} \beta_k \psi_k = B + \beta_{L-1} \psi_{L-1} + \beta_L \psi_L$$ (32)

Where $B$ is defined by equation(33).

D. Assessment Parameters

In order to get the systems’ assessment parameters, the balance equations ((17) to (34)) have been solved, using the spectral expansion method. Finally, the steady state transition probabilities, $\pi(i, j)$, and used for obtaining of the availability, reliability, and performability measures. The mean queue length ($MQL$) of calls which will be served according to the availability model, the new call blocking probability ($P_b$), and handoff forced termination probability ($P_{fh}$), can be evaluated as follows:

$$MQL = \sum_{j=0}^{L} \sum_{i=0}^{N} \pi(i, j) \quad (35)$$

$$P_b = \sum_{i=0}^{N} \sum_{j=0}^{L} \pi(i, j) \quad (36)$$

$$P_{fh} = \sum_{i=0}^{N} \pi(i, L) \quad (37)$$

IV. Numerical Results and Analysis

To investigate the effectiveness of the method presented and evaluate the performance of the proposed system, the current section is divided into three parts:
\[
B = -\beta_{L-1} \psi_{L-1} - \beta_L \psi_L - (\beta_{L-1} \psi_{L-1} y_{L-1} + \beta_L \psi_L y_L) (Q_1 + D^C)^{-1}
\]  

\[
V_{j-1} = \begin{cases} 
\frac{1}{B_j} (V_j [D^A + D^C - A]) ; & j = L \\
\frac{1}{B_j} (V_j [D^A_j + D^B_j + D^C_j - A_j] - V_{j+1} C_{j+1}) ; & 2 \leq j \leq (N - N_g) \\
\frac{1}{B_j} (V_j [D^A_j + D^B_j + D^C_j - A_j] - V_{j+1} C_{j+1}) ; & (N - N_g) < j \leq (L - 2) \\
\frac{V_j C_{j}}{[D^A_j + D^B_j - A_j]} ; & j = 1
\end{cases}
\]

A. Model Validation

In order to consider the validation of the obtained results, the system parameters have been chosen to match that of the previously published work in [3],[5],[13]. Performance of such system is shown for comparison purpose only. Figures 4 and 5, have the same parameters of [13], which may be summarized as follows: \(N = 10\), \(N_g = 0\) (i.e., without handoff priority), \(H = 10\), \(E(T_c) = 60\) sec, \(\eta = 2\) hr\(^{-1}\), \(\xi = 0.001\) hr\(^{-1}\), and \(L = N + H = 20\).

Figure 4, clearly shows that the obtained MQL is approximately identical with the presented results in [13]. On the other hand, figure 5 shows that the new call blocking probability has the same behavior of the results that have been presented in [13].

The main advantage of the developed model is the deployment of handoff priority scheme as well as the system’s ability for multi recovery. So, it may be concluded that: whereas the present model will maintain the same accuracy of the previously presented work at [7],[9],[13], it will introduce the influence of more realization scenarios.

B. Performability Results for Different Guard Channels

In order to investigate the effect of the number of guard channels \(N_g\) for different system capacities \(N\), the system assessment is done with the following chosen system parameters. These parameters are matched with [13],[15],[19] but with adding the process of pre-reserved guard channels. The parameters are as follows: \(N = 10\) to \(80\) (in order to investigate the different radio access technologies (RATs)), \(H = 10\), \(\mu_c = 0.0083\) sec\(^{-1}\), \(\eta = 0.6\) hr\(^{-1}\), \(\xi = 0.006\) hr\(^{-1}\), and \(L = N + H, N_g\) from 0 (i.e. no handoff priority is considered), up to 50 % (to represent the extreme case for handoff priority). The handoff arrival rate is chosen to represent 25% of the new call arrival rate. The arrival rate is chosen to have 80% of its maximum rate that is bounded by (2). The handoff request rate is function of the ratio between
the average velocity and the cell radius. This ratio is chosen to represent the case of system roll out in the urban areas. Figure (6), shows the mean queue length $MQL$ versus the number of guard channels.

Figure (6). $MQL$ vs $N_g$ for different system capacities

As illustrated in figure (6), it is shown that as $N_g$ increases, $MQL$ will decrease considerably. This occurs as a result of the system tendency to serve more calls/services as well as the number of guard channels increases. It is clearly evident that when the system capacity $L$ increases, $MQL$ will increase monotonically. This is because of the system may have more service requests and hence more contention opportunities.

Figure (7) shows the obtained results of the new call blocking probability $P_b$ versus the number of guard channels.

As illustrated in figure (7), it is shown that as $N_g$ increases, $P_b$ will increase considerably. This occurs as a result of the lack of the allowable space for serving new call traffic. It is clearly evident that when the system capacity $N$ increases, $P_b$ will decrease monotonically. This is because of the system may have more allowable space to be used for new service requests and hence less contention opportunities.

Figure (8) shows the obtained results of the new call blocking probability $P_{fh}$ versus the number of guard channels.

As illustrated in figure (8), it is shown that as $N_g$ increases, $P_{fh}$ will decrease considerably. This occurs as a result of the enlargements of the allowable space for serving handoff call traffic. It is clearly evident that when the system capacity $N$ increases, $P_{fh}$ will shrink monotonically. This is because of the system may have wider spaces to be used for handoff service requests and hence fewer contention opportunities.

C. Performability Results for Different Arrival Rates

The current subsection aims to investigate the effect of having different traffic loads. The system has been assessed for different new call arrival rates up to its maximum of arrival rate, (which is determined by means of (2)). The other system parameters are as follows: $N_g$ is taken at 20 % and 40% (as a practical considerations)of the total system capacity as mentioned in the previous subsection. $N$ varies from 20 to 80 channels, $H=10$, $\mu_c = 0.0083 sec^{-1}$, $\eta = 0.6 hr^{-1}$, $\xi = 0.006 hr^{-1}$, and $L = N + H$. The handoff arrival rate is chosen to represent 25% of the new call arrival rate. The average velocity to the cell radius ratio is chosen to represent
the case of network deployment in the urban areas. Figure (9) shows the obtained results of $MQL$ versus the new call arrival rate $\lambda_n$ for different system capacities.

![Fig. 9. $MQL$ vs New Call Arrival Rate for different system capacities](image)

As illustrated in figure (9), it is shown that as $\lambda_n$ increases, $MQL$ will increase considerably. This occurs as a result of the enlargements of the service requests over the same system capacity. It is clearly evident that when the system capacity $N$ increases, $MQL$ will be increased monotonically. This is because of the system may have more and more contention and so more system failures will be occurred. This reflects more extent for the serving queue length.

Figure (10) shows the obtained results of $P_b$ versus the new call arrival rate $\lambda_n$ for different system capacities.

![Fig. 10. $P_b$ vs New Call Arrival Rate for different system capacities](image)

As illustrated in figure (10), it is shown that as $\lambda_n$ increases, $P_b$ will increase heavily. This occurs as a result of the lack of the allowable system resources'. It is clearly evident that when the system capacity $N$ increases, $P_b$ will decrease monotonically. This is because of the system may have more opportunity to accept higher loads.

Figure (11) shows the obtained results of $P_{fh}$ versus the new call arrival rate $\lambda_n$ for different system capacities.

![Fig. 11. $P_{fh}$ vs New Call Arrival Rate for different system capacities](image)

As illustrated in figure (11), it is shown that as $\lambda_n$ increases, $P_{fh}$ will increase. This occurs as a result of the excessive contention at certain system capacity. It is clearly evident that when the system capacity $N$ increases, $P_{fh}$ will decrease extensively. This is because the system will have more opportunities to accept heavy loads.

V. CONCLUSION AND FUTURE WORK

This paper evaluates the performability of different RATs by means of considering not only teletraffic Markov models, but also, its availability. Unlike the previous work, an exact model and solution approach to evaluate the system’s performability. The handoff priority issues are considered as well as the multi breakdowns, and repairs. Spectral expansion method is used to derive the state transitions probabilities for a 2D Markov chain with QBD process. Numerical results are presented comparatively with results obtained by using the Markov reward rate approach in [13] for various performability measures. Numerical results have been obtained and presented
for various performability parameters, such as $MQL$, $P_b$, and $P_{fh}$. These parameters have been evaluated for different number of $N_g$ with different system capacities. It has been noticed that as $N_g$ increases, both of $MQL$ and $P_{fh}$ will decrease, whereas $P_b$ increases. This occurs as a result of increasing the available system’s resources for the handoff traffic. From the system capacity point of view at certain ratio of $N_g$, both of $P_b$ and $P_{fh}$ will decrease, whereas $MQL$ increases. This occurs as a result of reducing the opportunity of blocking or terminating calls. At the same time, this capacity increase will affect the system availability to recover its failures, so the $MQL$ will increase. Finally, the effect of varying of the incoming traffic loads has been investigated. The obtained results show that as $n$ increases, $MQL$, $P_b$, and $P_{fh}$ will increase. This occurs as a result of the excessive contention at certain system capacity. From the system capacity point of view at certain ratio of $N_g$, both of $P_b$ and $P_{fh}$ will decrease, whereas $MQL$ increases. This occurs as a result of the system’s ability to accommodate larger number of calls. On the other hand, it is clearly evident that when, the system capacity increases, $MQL$ will increase due to the system will be more susceptible to more and more breakdowns and so more queue length for repairing of these failures.

As a future direction, the presented work may extends its methodology for other systems with different class of service. Another possible extension to this work is the exploration of heterogeneous wireless networks and its availability models.

REFERENCES


Hesham ElBadawy Member in the IEEE Society for Vehicular Technology. Dr Hesham is an Associate Prof. and Head of Network Planning Dept Executive Manager for Mobile BTS Auditing Project in National Telecommunication Institute, Cairo, Egypt. Currently he is interested in LATEX based articles. Nowadays, he is doing a lot of works with himself to avoid his own collapse. I know that is may be considered crazy to say such things but we have to fight against every thing if we need to survive. In addition, I have a lot of papers concerning the performance issues, system modeling, queueing networks, mathematical based systems and so, I have a great chance to be promoted to the full professor class as soon as possible. On the other hand I have to perform some modifications for the current paper in order to get the IEEE approval INSHAALAH. email:heshamelbadawy@ieee.org.