OPSEARCH 46(3):289–302

THEORY AND METHODOLOGY

# Call processing delay analysis in cellular networks: a queuing model approach

**Vaneeta Jindal, S.Dharmaraja**

*Department of Mathematics, Indian Institute of Technology, Delhi, India*
*dharmar@maths.iitd.ac.in*

**Abstract**

Mobile devices generates a call request message while initiating a call, which is then sent to a base station (BS). A BS processes a call request message and then takes a decision of acceptance or rejection for the call. It is important to analyze the total time, which includes the waiting time and the processing time, spent by a call request message in the system. For if, this time is greater than the permissible delay, call will be blocked. Hence, the quality of service is degraded which is not acceptable to the service providers. Further, failures (hardware and software related) and their recovery increase the delay. With this background, in this paper, we present queuing models for the analysis of delay experienced by a call request message for two cases: first, services of BS are not interrupted and second, services are interrupted due to the occurrence of failures at the BS. In the end, we discuss special cases of proposed queuing models based on the distribution of the processing times.

**Keywords**

Call processing, Delay, Queuing model, M/G/1 queues, Service interruption

## 1. Introduction

Cellular networks divide a geographic area into smaller regions called cells. A base station (BS), which provides wireless connectivity through wireless channels, serves each cell. Several BSs are connected to a base station controller (BSC), which are then connected to a network subsystem. The network subsystem consists of mobile switching center (MSC), home location register (HLR) and visitor location register (VLR). MSC is responsible for authenticating the mobile user, storing location information and routing user's calls to appropriate networks.

Figure 1 shows the generic architecture of a cellular network. To establish a communication session or a call, the mobile station (MS) sends a request

Received: 4 February 2008 / Accepted: 4 August 2008
© Operational Research Society of India, 2009

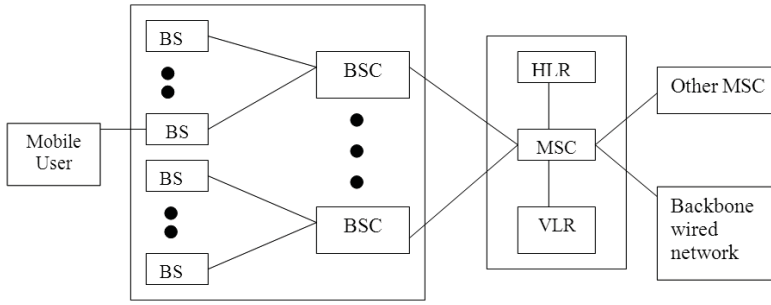


**Fig. 1.** Architecture of cellular networks

for radio channels through a channel access procedure. On receiving call request messages successfully, BS interacts with its BSC and the network subsystem to carry out the processing of the call request message and hence takes the decision to accept or reject the call. A BS accepts the call if an idle wireless channel can be allocated to the call for the communication. To utilize the limited radio resources at BS efficiently, call admission control (CAC) schemes are incorporated. Some of the common CAC schemes include fixed guard channel scheme proposed by Haring et al. [1], dynamic CAC scheme proposed by Li et al. [2] and channel borrowing CAC scheme, Cao et al. [3]. Ahmed [4] gives a comprehensive survey on CAC schemes.

The processing time of a call request message varies for different CAC schemes. For example, in fixed guard channel scheme, a BS searches for an idle channel in its channel pool. If it finds an idle channel, this channel is allocated to the call request. In channel borrowing scheme, if a BS does not possess an idle channel, it borrows a channel from one of its neighbor cells and allocate this channel to the call request. It is apparent that the time taken for processing of a call request message in both the schemes is not same and is random.

From the view point of service providers, it is necessary to analyze the delay in processing of a call request message and its waiting time in a queue. For if, this time is greater than the permissible delay, the corresponding call will be blocked. Hence, the Quality of Service (QoS), is degraded, which is unacceptable.

Tipper et al. [5] discussed that the failures at BS increase call processing time and degrade QoS. Chen et al. [6] discussed about the hardware failures (such as failures in electro-mechanical equipments) and software failures

(such as Heisenbugs and Bohrbugs). To mitigate the impact of the failures at BS, several fault tolerant strategies are incorporated by the network designers. Varshney et al. [7] proposed fault tolerance strategies that are responsible for recovery of failures at a BS. After the recovery, BS presumes processing of a call request message. Thus, failures and their recovery delay the processing of a call request message. It becomes important to compute mean processing time of a call request message in the incidence of failures and their recovery. In this paper, we present a queuing model to analyze the processing delay of a call request. The failures at BS negatively affect the call blocking and dropping probabilities. The call processing delay is further increased due to the failures at a BS. Fault tolerance strategies are then applied to restore the services at BS. Before incorporating fault tolerance strategies it is essential to analyze overhead incurred by them. Hence, in the presence of failures and their recovery, we model the processing delay of call request messages by a $M/G/1$ queuing model with vacation, where vacation is due to failures and services of a BS are resumed after recovery of the failures.

The rest of this paper is organized as follows: In Section 2, we describe the system model of a cellular network. In Sections 3, we present the analytical model to compute the mean and variance of the total time spent in the system by a call request message for the case when service is not interrupted by the failures at BS. In Section 4, similar analysis is carried out by considering the failures and recovery at BS. In Section 5, we obtain mean and variance of the total time spent in the system for specific service distributions. Finally, in Section 6, we discuss the inferences drawn from these models.

## 2. System model

We consider the future generation cellular networks; the network traffic consists of real time service (RTS) calls (for example, voice and video conference) and non-real time service (NRTS) calls (for example, SMS, data transfer and e-mail). We assume that RTS and NRTS call request messages arrive independently according to Poisson process with rates $\lambda_1$ and $\lambda_2$, respectively. Both types of call request messages join a single queue. Therefore, the arrival of call request messages at a single queue is a Poisson process with rate $\lambda = \lambda_1 + \lambda_2$. Each call request message is processed on first-come-first serve basis. Processing time of a call request message depends on the CAC scheme employed at a BS. After the processing, the decision is taken to accept or reject a call. We assume an infinite buffer for call request messages. This assumption is justified because only after the processing of a call request message, the call is accepted or rejected.

The service time, denoted by $S$, of a BS is defined as the time duration taken by a BS for processing of a call request message. The distribution of $S$ depends on a CAC scheme. For example, in channel borrowing scheme by Cao et al. [3], the time to process call request message is sum of the time taken to search for an idle channel in the channel pool and the time to borrow an idle channel from one of its neighbor cells. This channel borrowing process involves exchanges of control messages between the BS and its neighbors. The processing time then includes propagation time, transmission time and the time for computing the set of channels that can be borrowed from the neighboring cells. Thus, the assumption of standard exponential distribution for the processing time of a call request message is not valid. We assume that the service times of RTS and NRTS call request messages are independent and identically distributed with cumulative distribution function $G(\cdot)$ and mean $1/\mu$. Note that the call holding times for RTS and NRTS calls are possibly different.

Failures at a BS further delays the processing of a call request message. Fault tolerance strategies are incorporated for handling these failures. The overhead associated with these strategies is the recovery time of the failures at BS. It is essential to analyze the impact on call processing delay for call request messages in the presence of failures and their recovery. For these purpose, we construct a queuing model. In next section, we first develop a $M/G/1$ queuing model, without considering the failures at BS.

## 3. Queuing model for delay in processing without service interruptions

We begin by describing the analytical model for computing the mean and variance of the total time spent in the system by a call request message. This total time accounts for the processing delay of a call request message. Let the random variable $X(t)$ represents the number of call request messages in the BS at any time $t$ and $\{X(t), t \geq 0\}$ is a stochastic process with state space $\{0,1,2,\ldots\}$. We assume that the service time of a BS is generally distributed and the network traffic is following Poisson process. Further, we consider that no failures occur at BS. With these assumptions, $\{X(t), t \geq 0\}$ is a non-Markovian process and is modeled as a $M/G/1$ queuing system. Figure 1 shows the queuing model for the system without service interruptions. The service discipline is first-come-first-serve and the inter-arrival times and service times are independent.

Let $M_n$ be the number of call requests in the system at the departure of a $n^{th}$ call request. Then,

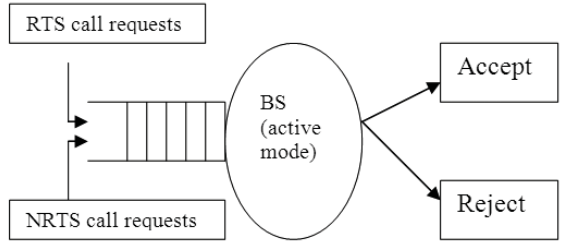


**Fig. 2.** A queuing model without service interruption at the BS

$$M_{n+1} = \begin{cases} M_n - 1 + A_{n+1}, & M_n \geq 1 \\ A_{n+1}, & M_n \geq 0, \end{cases}$$

where $A_{n+1}$ is the number of call requests that arrive during service time of a $n + 1^{th}$ call request. Then, $\{M_n, n = 1, 2...\}$ is an embedded discrete time Markov chain for $\{X(t), t \geq 0\}$. Each random variable $A_{n+1}$ is independent and identically distributed.

Let $\rho = \lambda/\mu$ , where $\lambda$ is the mean arrival rate of call request messages and $1/\mu$ is the mean service time of a BS. It has been shown in [8] that $M/G/1$ embedded chain $\{M_n, n = 1, 2...\}$ is an irreducible, positive recurrent and aperiodic when $\rho < 1$ and hence it posses a long run distribution under this condition. Since PASTA (Poisson arrival sees time averages) theorem holds for $M/G/1$ queuing model, the steady state solution for $\{M_n, n = 1, 2...\}$ is same for $\{X(t), t \geq 0\}$.

**Mean total time spent by a call request message**

Let $N$ and $T_1$ denote the steady state system size and the total time spent in the system respectively. Then, by Pollaczek-Khintchine formula for $M/G/1$ queuing system [8] expected system size is given as:

$$E[N] = \rho + (\rho^2 + \lambda^2 \sigma_S^2) / (2(1-\rho)) \tag{2}$$

where $\sigma_S^2$ is variance of the service time distribution.

Let $Q$ and $W_1$ be the steady state queue size and the waiting time in the queue respectively. Then, using Equation (2), Little's formula $E[Q] = \lambda E[W_1]$ and the relations $E[T_1] = E[W_1] + 1/\mu$ and $E[Q] = E[N] - \lambda/\mu$ , we obtain expected total time spent by a call request message in the system as

$$E[T_1] = \left.\lambda^2\sigma_S^2 + (2\rho - \rho^2)\right/(2\lambda(1-\rho)) \tag{3}$$

**Variance of total time spent by a call request message**

We next compute variance of the total time spent by a call request message by differentiating twice Laplace-Stieltjes Transform (LST) of the cumulative distribution function $B_1(t)$ for $W_1$. Note Note that the LST of a function $F(t)$, denoted by $\tilde{F}(s)$ is given as

$$\tilde{F}(s) = \int_0^\infty e^{-st} dF(t).$$

We know that, if $F(t)$ is a cumulative distribution function of a random variable $X$, then $n^{th}$ order moment of $X$, denoted by $E[X^n]$, is obtained by taking $n^{th}$ derivative of $\tilde{F}(s)$ and then evaluating at $s = 0$.

Let $\tilde{B}_1(s)$ be the LST of $B_1(t)$. It is shown that [8]:

$$\tilde{B}_1(s) = \left.((1-\rho)s)\right/(s - \lambda[1 - \tilde{G}(s)]) \tag{4}$$

where $\tilde{G}(s)$ is the LST of $G(t)$, the cumulative distribution function of the service time.

By differentiating Equation (4), the first and second order moments of $W_1$ are given as

$$E[W_1] = \left.(\rho^2 + \lambda^2\sigma_S^2)\right/(2(1-\rho)\lambda) \tag{5}$$

$$E[W_1^2] = \left.(3E[W_1](\rho^2 + \lambda^2\sigma_S^2) + E[S^3]\lambda^2)\right/(3\lambda(1-\rho)) \tag{6}$$

where $E[S^3]$ is the third order moment of service time and $\sigma_S^2$ is its variance. Since total time spent by a call request message is sum of the waiting time and the service time, we have $T_1 = W_1 + S$. Then, the following relation gives its variance:

$$Var(T_1) = Var(W_1) + Var(S). \tag{7}$$

Using relation $Var(W_1) = E[W_1^2] - (E[W_1])^2$ and substituting for $E[W_1]$ and $E[W_1^2]$ from Equations (5) and (6) respectively, we find that

$$Var(W_1) = \left.((\lambda^2\sigma_S^2 + \rho^2)^2)\right/(2(1-\rho)^2 + \lambda^2) + \left.(\lambda^2 E[S^3])\right/(3(1-\rho)\lambda). \tag{8}$$

Finally, substituting for $Var(W_1)$ in Equation (7), variance of time spent in the system by call request message is given by

$$Var(T_1) = \frac{((\lambda^2\sigma_S^2 + \rho^2)^2)}{(2(1-\rho)^2 + \lambda^2)} + \frac{(\lambda^2 E[S^3])}{(3(1-\rho)\lambda)} + Var(S) \ . \quad (9)$$

## 4. Queuing model for delay in call processing with service interruptions

The failures (hardware and software) and their recovery interrupt the services at a BS and hence delay the processing of a call request message of RTS and NRTS types. In this section, we analyze how service interruptions affect the total time spent in the system by a call request message. The system is modeled as a $M/G/1$ queuing system with vacations where the vacations are due to the service interruptions. The corresponding queuing model is shown in Figure 2

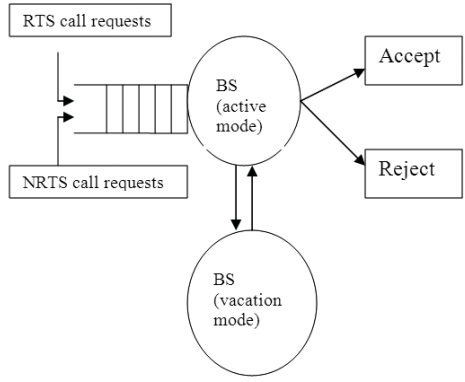


**Fig. 2.** A queuing model with service interruptions at the BS

RTS and NRTS call request messages are assumed to be arriving independently at a BS according to Poisson process with rates $\lambda_1$ and $\lambda_2$ respectively, so that arrival rate of call request messages is $\lambda = \lambda_1 + \lambda_2$. Service time is independent and identically distributed with cumulative distribution function $G(\cdot)$. We assume time to failures to be exponentially distributed with parameter $\alpha$. The time to repair is the duration of an interruption and is assumed to be generally (non-exponential) distributed with cumulative distribution function $V(\cdot)$.

The service interruptions can be non-preemptive or preemptive. In a preemptive service interruption, the service of the call request message is stopped immediately with the occurrence of the failures. On the other hand, non-preemptive interruptions do not affect the ongoing service of a call request message until the end of the service period. We assume non-preemptive service interruptions in this model. Further, we assume that service is interrupted only after 'n' number of failures. Before the beginning of the next service, these failures have to be recovered. We assume that no further failures occur during the recovery of these failures. Gaver [9] had considered other types of service interruptions such as preemptive-repeat-identical interruptions and preemptive-repeat-different interruptions. Therefore, it is necessary to consider an appropriate type of service interruption depending on the system. Gaver [10] studied the effects of interruptions on the distributions of important measures such as busy period, queue length and waiting time for a $M/G/1$ queuing model.

Due to the interruptions, there is a break in the service of the ongoing call request message. After a short duration (which is the recovery time of the failures), the service is resumed. In order to combine the effect of duration of interruptions and service time, we introduce the notion of completion time. The completion time is defined as the time that elapses from the instant the processing of a call request begins until the time the next call request message processing begins. In the absence of interruption due to failures, the completion time is same as the service time. In this model, we consider the non-preemptive interruption. Using the results reported by Gaver [9] and [10], the expected total time spent in the system is obtained.

**Mean total time spent by a call request message**

Let $C_j$ denote the completion time of a $j^{th}$ call request, $S_j$ is the service time, $D(i)$ is the duration of an $i^{th}$ interruption that occurs during the service period $S_j$ and $n$ is the number of failures that have occurred at the BS. Operation and maintenance center of the network system keep track of the number and type of failures occurring at a BS. Then

$$C_j = S_j + \sum_{i=0}^{n} D(i) \cdot \tag{10}$$

Since the completion times $C_j$ and the service times $S_j$ are independent and identically distributed for each $j$, we denote each $C_j$ by $C$ and $S_j$ by $S$. Then Equation (10) is re-written as:

$$C = S + \sum_{i=0}^{n} D(i). \tag{11}$$

Let $H(\cdot)$ be the cumulative distribution function of $C$ and $\tilde{H}(s)$ be its LST. Following the procedure in [10], $\tilde{H}(s)$ is obtained as:

$$\tilde{H}(s) = \tilde{G}(s + \alpha - \alpha\tilde{V}(s)), \tag{12}$$

where $\alpha$ is the rate of occurrence of interruptions and $\tilde{V}(s)$ and $\tilde{G}(s)$ denote LST of the $V(\cdot)$ and $G(\cdot)$ respectively. The LST of the completion time will then be used to compute the LST of waiting time of a call request message in the system.

Let $T_2$ denotes total time of a call request message in the system with the cumulative distribution function $B_2(t)$ . Then, following the methodology of Gaver [9], the expression for LST of $B_2(t)$ is obtained as

$$\tilde{B}_2(s) = \frac{((1-\rho)s)}{(s - \lambda + \lambda\tilde{H}(s))} \tag{13}$$

where $\tilde{H}(s)$ is given by Equation (12). Differentiating $\tilde{B}_2(s)$ with respect to $s$ and evaluating at $s = 0$, mean total time spent in the system is given by following equation $s = 0$ mean total time spent in the system is given by following equation:

$$E[T_2] = \frac{\rho}{2(1-\rho)}\left[1 + \frac{\sigma_C^2}{(E[C])^2}\right]E[C], \tag{14}$$

where $E[C]$ and $\sigma_C^2$ are, respectively, the mean and variance of the completion time $C$ and $\rho = \lambda/\mu$ as defined in Section 2.

**Variance of total time spent by a call request message**

The variance of $T_2$ is given as:

$$Var(T_2) = \left\{\frac{(\rho E[C^2])}{(2E[C](1-\rho))}\right\}^2 + \frac{(\rho E[C^3])}{(3(1-\rho)E[C])}, \tag{15}$$

where $E[C^k]$ is the kth order moments of the completion time $C$ and is obtained by differentiating $\tilde{H}(s)$ with respect to $s$ and evaluating at $s = 0$ . From the values of mean and variance, we can investigate how the random values of the time spent in the system will differ from the mean time spent in the system .

## 5. Special cases

In this section, we derive mean and variance of the delay in processing of call request messages for some specific distributions of service times of

a BS. We assume that the mean service time is $1/\mu$ for all the following three cases.

**Case I: Exponentially distributed service times**

We consider the service time $S$, to be exponentially distributed. First, we derive the mean and variance of the delay with an uninterrupted service at BS. Then, the cumulative distribution function $G(t)$ of is $(1-e^{-\mu t})$ and

$\tilde{G}(s) = \mu/(\mu + s)$.

Note that the $n^{th}$ order moments exist for exponential distribution. The first, second and third order moments are respectively given as

$$E[S] = 1/\mu, E[S^3] = 2/\mu^2, E[S^3] = 6/\mu^3,$$

and variance $\sigma_S^2$ is $1/\mu^2$. Note that $\rho = \lambda E[S] = \lambda/\mu$. Substituting for these above values in Equations (3) and (9), we obtain,

$$E[T_1] = \mu/(\mu - \lambda), Var(T_1) = (1 + 2\lambda\mu)/(\mu(\mu - \lambda))^2.$$

Note that the mean time spent in the system for this case is same as the mean time spent in the system for a $M/M/1$ queuing model [8].

Next, we consider the case of interrupted services where the interruptions are caused due to the failures at the BS. In this case, the service time $S$ in above is replaced by the completion time $C$ that includes the time to failures and recovery in addition to service time. We assume that the time to failure is exponentially distributed with parameter $\alpha$. The time to recovery is generally (non-exponentially) distributed with cumulative distribution function $V(t)$. However, for the computation purpose, we assume exponential distribution for the time to recovery with LST $\tilde{V}(s) = \beta/(\beta + s)$. The LST of distribution function of $C$ is then given as

$$\tilde{H}(s) = (\mu(\beta + s))/(s^2 + s(\alpha + \beta + \mu) + \mu\beta).$$

Differentiating $\tilde{H}(s)$ with respect to $s$ and evaluating at $s = 0$, we obtain the first, second and third order moments as follows:

$$E[C] = (\alpha + \beta)/\mu\beta, \quad E[C^2] = (2[(\alpha + \beta)^2 + \mu\alpha])/(\mu\beta)^2,$$

$$E[C^3] = (2[(\alpha+\beta+\mu)^2][3(\alpha+\beta)^2 + 3\mu\alpha - \mu\beta]) \Big/ (\mu\beta)^3 .$$

Substituting for these above values in Equations (14) and (15), we obtain the mean and variance of the total time spent by a call request message for the case when the service at BS is interrupted by the failures.

As expected, the time spent in the system with interrupted services is longer than the corresponding time spent in the system without interruptions. This observation is implied by a consideration of realistic systems.

**Case II: Deterministic service times**

We next consider the case when the BS takes the constant time, say $1/\mu$ for processing of call request messages. The cumulative distribution functions for $S$ and its LST are respectively given as

$$G(t) = \begin{cases} 0, & t < 1/\mu \\ 1, & t \geq 1/\mu \end{cases}, \quad \tilde{G}(s) = e^{-s/\mu} \Big/ s$$

The mean and variance of the total time spent by a call request message are then obtained from Equations (3) and (9).

To obtain the mean and variance of the total time by a call request message spent in the system, we consider the completion time $C$ instead of the service time $S$. The LST of cumulative distribution function $H(t)$ of $C$ is given as:

$$\tilde{H}(s) = \tilde{G}(s+\alpha-\alpha\tilde{V}(s)) = (e^{-(s+\alpha-\alpha\tilde{V}(s))^k}) \Big/ (s+\alpha-\alpha\tilde{V}(s)),$$

Where $\tilde{V}(s) = \beta/(\beta+s)$ is the LST of the cumulative distribution function of the exponentially distributed recovery times.

The $n^{th}(n=1,2,3)$ order moments of the completion time $C$ are obtained by taking the $n^{th}$ derivative of $\tilde{H}(s)$ with respect to $s$ and evaluating at $s=0$. Then, substituting for these values accordingly in Equations (14) and (15), we obtain mean and variance of the total time spent by a call request message.

**Case III: Erlang of type k distribution for service times**

We now consider the case of Erlang of type $k$ distribution for service times at the BS. The Erlang of type $k$ distribution for service times implies that the service at BS consists of $k$ steps (or phases) such that the time spent in each of the phase is exponentially distributed with rate $k\mu$ and all steps are independent and identical .

The LST of $G(t)$ , the cumulative distribution function of $S$ is given as:

$$\tilde{G}(s) = \left( \mu / (\mu + s) \right)^k.$$

The mean and variance of the total time spent by a call request message are then obtained from Equations (3) and (9) for the case of uninterrupted service at BS.

Next, to compute the mean and variance of the total time spent by a call request message in the system with interrupted service, the LST for the $H(t)$, the cumulative distribution function of completion time $C$ is obtained as:

$$\tilde{H}(s) = \left( (\mu(\beta + s)) / (s^2 + s(\alpha + \beta + \mu) + \mu\beta) \right)^k$$

By taking the $n^{th}$ derivative of $\tilde{H}(s)$ with respect to $s$ and evaluating it at $s = 0$, we obtain $n^{th}$ order moments of $C$. Then, substituting for these values accordingly in Equations (14) and (15), we obtain the mean and variance of the total time spent by a call request message.

## 6. Discussion and future work

A call request message is sent to a BS by a MS that wants to initiate a call. In this paper, we analyze the total time spent by a call request message at the BS. The time duration includes the waiting time in the queue and the time for processing of the call request message. The processing time depends on the CAC scheme incorporated in the system. The failures and their recovery procedures interrupt the service at BS and hence delay the processing of a call request message. This, in turn, increase the total time spent by a call request message in the system. Both the cases, uninterrupted and interrupted services were considered to obtain the mean and variance of the total time spent by a call request message in the system.

This analysis is important to the network designers and service providers in the following ways:

1. After receiving a call request message, BS starts its processing. The processing time depends on the CAC scheme implemented at the network. The network designers can choose an efficient CAC scheme to meet their system requirements and QoS of the network traffic. For example, the system designers may choose a CAC scheme corresponding to which BS has less processing time, for the network that constitute mainly of delay sensitive calls.
2. It is apparent that the time spent in the system with interrupted services depends on the recovery time. The network designers can incorporate the fault tolerant strategies that take less time to recovery for failures and hence reduce the duration of an interruption. For the purpose of illustration, we had considered the recovery time to be exponentially distributed and then computed the mean and variance of the time spent in the system. However, one can further analyze the mean and variance of the time spent in the system for non-exponential distributions of the recovery time as well.
3. It is an implied fact that with an increase in the number of call request messages, the delay also increases. The mean and variance of the time spent by a call request message in the system (with an infinite buffer for call request messages) can be used to account for the permissible delay. This permissible delay depends on the CAC scheme implemented at the networks. If the permissible delay exceeds the delay tolerance of the calls in the steady state, the network designers can modify the network design to operate as a system with a finite buffer for call request message. The buffer size can be optimized subject to the constraint that the permissible delay is less than the delay tolerance of the calls.

## Acknowledgments

This research work was supported by Department of Science and Technology, India, under the grant number RP 1907. One of the authors (V.J.) would like to thank the CSIR India, for their financial support provided to her.

## References

[1] Haring G., Marie R., Puigjaner R., Trivedi K.S.: Loss formulas and their applications to optimization for cellular network, IEEE Trans Veh Tech, **50**(3), 664–673 (2001)

[2] Li B., Li L., Li B., Sivalingam K.M., Cao X.: Call admission control for

voice/data integrated cellular networks: Performance analysis and comparative Study, IEEE J Sel Area Comm, **22**(4), 706–718 (2004)

[3] Cao G., Singhal M.: Distributed fault tolerant channel allocation for cellular networks, IEEE J Sel Area Comm, **18**(7), 1326–1337 (2000)

[4] Ahmed M.: Call admission control in wireless networks: A comprehensive Study, IEEE Comm Surv Tuto, **7**(1), 50–69 (2005)

[5] Tipper D., Dahlberg T., Shin H., Charnsripriyo C.: Providing fault tolerance in wireless access networks, IEEE Comm Mag, **40**(1), 58–64 (2002)

[6] Chen D., Dharmaraja S. Chen D., Li L., Trivedi K.S., Some R.R., Nikora A.P.: Reliability and availability analysis of the JPL remote exploration experimentation system, proceedings international conference on dependable and system networks (DSN 2002), USA, 337–342 (2002)

[7] Varshney U., Malloy A.: Multilevel fault tolerance in infrastructure oriented wireless access networks: framework and performance evaluation, Int J Netw Manag, **16**(5), 351–374 (2006)

[8] Gross D., Harris C.M.: Fundamentals of queuing theory, John Wiley, Third Edition, New York, (1998)

[9] Gaver D.: A waiting-line with interrupted service, Including Priorities, J Roy Stat Soc, **B24**, 73–90 (1962)

[10] Gaver D.: Imbedded Markov chain analysis of a waiting-line process in continuous time, Ann Math Stat, **30**, 698–720 (1959)