System-Level Analysis of Mobile Cellular Networks Considering Link Unreliability

Carmen B. Rodríguez-Estrello, Student Member, IEEE, Genaro Hernández-Valdez, Member, IEEE, and Felipe A. Cruz-Pérez, Senior Member, IEEE

Abstract—Traditionally, system-level performance evaluation of mobile wireless communication networks has been addressed by only considering resource insufficiency, whereas the effect of an unreliable wireless channel has largely been ignored because of the complexity that its inclusion entails. To fill this void, a general analytical model for the system-level performance evaluation of mobile wireless networks taking into account both resource insufficiency and link unreliability is proposed in this paper. The effect of link unreliability is captured through the appropriate probabilistic characterization of the “unencumbered call-interruption time.” Additionally, useful functional relationships between the call-interruption processes of the proposed analytical model and some system-level parameters that can easily be obtained from statistics collected at base stations (BSs) are derived. For the sake of generality, the involved time variables (i.e., cell dwell time, unencumbered call-interruption time, and channel holding time) are considered generally distributed. Then, general and easily computable mathematical expressions for many useful performance metrics under more realistic considerations are obtained. The analytical model proposed here is able to provide new and important insights into the dependence of system performance on link unreliability. Such understanding of this teletraffic engineering issue is vital for planning, designing, dimensioning, and optimizing mobile cellular networks for present and future wireless communication systems beyond the third generation.

Index Terms—Call-dropping process, call forced-termination probability, cell dwell time, channel holding time, handoff failure, link unreliability, mobility, resource insufficiency, system-level performance evaluation, unencumbered call-interruption time, wireless cellular networks.

I. INTRODUCTION

I
T HAS WIDELY been accepted that call forced-termination probability is one of the most important QoS metrics for the performance evaluation of present and future mobile wireless networks [1]–[6]. In the context of packet-switched mobile communication networks, call forced-termination probability is particularly important in the performance evaluation of conversational and real-time services (i.e., voice, audio, music, videophone, videoconference, etc.) [4]–[6].

In mobile wireless networks, a call is forced to terminate because of two fundamental features, that is, resource insufficiency and link unreliability. When a mobile user moves into a different cell during the course of a call, a handoff must be performed. If no radio resources are available in the target cell, the call is said to be forced to terminate due to resource insufficiency. On the other hand, during the course of a call, the physical link between a BS and a mobile station (MS) may become severally degraded due to propagation impairments such as multipath fading, shadowing or path loss, and interference.1 Hence, the call may abnormally be terminated. In this case, the call is said to be forced to terminate due to link unreliability [7]–[30]. In general, the physical link is said to be unreliable if the experienced signal-to-interference ratio (SIR) is below than a minimum required value (SIR threshold) for more than a specified period of time (time threshold) [8]–[20]. In this research direction, the authors in [10]–[30] state that the reason for a large part of calls forced to terminate is the highly time varying and unreliable wireless channel. They conclude that it is fundamental to study system-level and link-level performances by taking into account the effect of wireless channel impairments in mobile networks. Nonetheless, the system-level performance evaluation of mobile wireless communication networks has commonly been addressed by only considering resource insufficiency, whereas link unreliability has received not enough attention because of the complexity that its inclusion entails [10]–[16]. Additionally, for convenience and mathematical tractability, system performance is commonly investigated under the assumption that both the unencumbered service time and the cell dwell time are exponentially distributed, that is, only the mean values of these time variables have been relevant for the performance metrics [31]–[34]. However, in many situations, these assumptions are not appropriate [35]–[51].

To fill this void, in this paper, an elegant and general analytical model for the system-level performance evaluation of mobile wireless networks taking into account both resource insufficiency and link unreliability is proposed and mathematically analyzed. The proposed mathematical approach is based on the use of a simple call-interruption process to model the effect of link unreliability through the introduction of the so-called “unencumbered call-interruption time.” Additionally,

1In CDMA-based systems, link unreliability is experimented due to the initial power allocation. The initial handoff decisions are made for individual connections independent of the other connections or the BS power availability. Therefore, current connections may suffer from an unreliable connection [39].
teletraffic analysis is developed considering that the involved time variables (i.e., cell dwell time, unencumbered call-interruption time, and channel holding time) are generally distributed random variables (RVs), that is, the commonly exponential assumption imposed on these time variables is relaxed. To appreciate the originality and effectiveness of our proposed analytical model, numerical results from concrete examples that connect our mathematical analysis to some practical cases are presented. The analytical model proposed here represents a step toward the development of a general and analytical tractable system-level modeling tool for the design of mobile wireless communication networks under more realistic considerations.

The remainder of this paper is structured as follows. A comprehensive overview of the previously published work on link unreliability is presented in Section II. In Section II, the relevance of nonexponential assumption on the probability distribution for the involved time variables is discussed. Section III describes the proposed model for the probabilistic characterization of the call-interruption due to link unreliability. In Section IV, a description of the considered system model is provided. Following this, the probability distribution function of the unencumbered call-interruption time is derived from the link- and system-level parameters in Sections V and VI, respectively. Novel mathematical expressions for several relevant performance metrics are obtained in Section VII. In Section VIII, a numerical investigation is carried out to discuss the impact of link unreliability on system performance. Finally, Section IX shows the conclusions and outlines the future work.

II. PREVIOUSLY PUBLISHED RELATED WORK

A. Link Unreliability

In a well-established cellular network and from the call forced-termination point of view, the handoff failure can usually be a negligible event\(^2\) [15]. On the contrary, the authors of [10] and [11] remark that the reasons for the large part of dropped calls are poor wireless channel conditions. In particular, the analysis of measured data traffic supplied by Vodafone Italy is a good example for these phenomena [15]. However, only relatively few recent studies have addressed the effect of link unreliability on the performance of mobile wireless communication networks [9]–[13], [27]–[30]. To support this, the authors of [30] textually state: “From the aforementioned studies, we are impressed that these studies focus upon the limited bandwidth effect upon the call or network performance, and hence, the lossy wireless link is ignored.” In this research direction, most of these studies have been performed only through link-level analysis, whereas the system-level analysis has not substantially been addressed due to its analytical complexity. System-level analysis involves variables such as channel holding times for successfully terminated calls and for forced terminated calls, which are easily obtained at real networks, whereas link-level statistics involve channel characterization (i.e., in terms of channel state characteristics and the probability distributions of the channel state duration), which, in general, are not easily obtained at real cellular networks.

Additionally, most of the work devoted to study the impact of link unreliability on system performance [9]–[11], [16], [23]–[25], [27] has considered a link-level channel model to study how the channel impairments affect the system performance. The Gilbert–Elliot [52], [53] and Fritchman [54] channel models have widely been used for this purpose. For instance, the authors of [16] proposed a queuing system with impaired wireless channel based on the Markov chain approach assuming that the unreliable wireless channel can be modeled by the Gilbert–Elliot channel. Analysis is performed by assuming that the wireless channel is the server. Then, the service rate becomes time varying due to propagation impairments. A model to quantify the performance of a queue with respect to such an impaired wireless channel is then developed. In a similar work [27], the Gilbert–Elliot channel model is proposed for the downlink performance evaluation of wireless code-division multiple-access (WCDMA) cellular systems.

In addition, in [27], the Gilbert–Elliot channel parameters in terms of the mean fade and nonfade durations are obtained. The authors of [23]–[25] introduced new QoS performance metrics to study the impact of channel impairments on the system performance. The proposed QoS performance metrics are based on the satisfied user criteria recommended by the Universal Mobile Telecommunications System (UMTS) (i.e., satisfied user probability for speech services and satisfied user probability for data services). Specifically, [25] deals with the UMTS QoS recommendation for packet-based networks. In [25], an analytical model is developed to examine both the packet loss rate and the time distribution of lost packets of N-multiplexed voice sources in 3G wireless networks. The authors of [24] propose a composite performance index (called satisfied user probability) based on the dropped-call probability and the session outage percentage due to link unreliability in wireless communication networks. This performance index is calculated by considering a Gilbert–Elliot channel model with negative exponentially distributed state durations. However, none of the preceding papers (i.e., [16], [23]–[25], and [27]) consider the users’ mobility. Consequently, call forced-termination due to resource insufficiency is not addressed.

Other related papers devoted to study the impact of link unreliability on system performance are [9]–[13], and [30], which use either the Gilbert–Elliot or Fritchman model to characterize the time-variant wireless channel. In [9]–[13] and [30], Zhang et al. derived mathematical expressions for the probability that a call is successfully completed considering the concurrent impacts of bad quality in the channel and the lack of radio resources. In a more recent work [14], Zhang studied the impact of Rayleigh fast fading on various teletraffic QoS metrics in wireless networks (i.e., channel holding time, handoff probability, handoff call-arrival rate, call-blocking probability, call-completion probability, and call forced-termination probability) by taking into account the carrier frequency, maximum Doppler frequency, and fade margin. In [30], a closed-form formula for the call-completion probability is developed under the generalized wireless channel model and general call holding time distribution using the theory of complex variables.

\(^2\)Furthermore, in CDMA-based cellular systems, some proposed CAC strategies consider call dropping to mitigate the outage degradation [39].
and transform techniques (Laplace–Stieltjes transform and z-transform). From the teletraffic point of view, a system-level-based modeling of the link unreliability is preferred over link-level-based modeling because fewer state variables are needed. However, the mathematical models considered in [9]–[14] and [30] are based on the link-level statistics, which are not easily obtained by direct measures. Contrary to Zhang’s works, in this paper, the effect of link unreliability is captured through easily obtained system-level quantities that allow directly including the effect of link unreliability on the channel occupancy in the teletraffic analysis.

Only a few recently published studies have addressed the system-level analysis considering the link unreliability for wireless networks [20], [28], [29]. In [20], a queuing model to evaluate the performance of code-division multiple-access (CDMA) reverse link in a multiple-cell scenario was developed. In that work, a quasi-birth-and-death process was used to capture the variation of traffic loads in cells. Then, the authors of [20] obtained the stationary distribution of the system and some performance indicators, such as the outage probability of existing calls, blocking probability of new calls, average carried traffic in a cell, and dropping frequency of ongoing calls. Nonetheless, the mobility of mobile users and soft handoff are not modeled in [20]. Moreover, the system performance is studied assuming that the call dwell time is exponential distributed. Recently, a teletraffic model was proposed to evaluate the performance of time-division multiple-access (TDMA)-based [28] and CDMA-based [29] cellular networks considering both resource insufficiency and link unreliability. In [28] and [29], the effect of link unreliability is captured by an interruption Poisson process, which is characterized by the mean time of the “unencumbered call-interruption time.” This interruption process is characterized via system-level statistics based on the channel holding time, which is easily measured at BSs. The main drawback of [28] and [29] is that the involved times are considered negative exponentially distributed RVs. However, the next section shows that in many situations these assumptions are neither appropriate nor realistic [37], [38], [41]–[51].

B. Nonexponential Assumptions for Involved Times

This section provides a detailed overview of studies on the nonexponential assumptions for cell dwell time, unencumbered call-interruption time, and channel holding time. These works remark the necessity of considering the involved time variables as general distributed RVs. Moreover, emphasis is done on the use of phase-type distributions due to their mathematically tractability. The channel holding time, cell dwell time, and unencumbered service time are three fundamental time variables to model and analyze the performance of mobile wireless networks. These time variables are of primary importance to derive other key performance metrics such as handoff probability, handoff rate, new call-blocking probability, and call forced-termination probability. Typically, for convenience and tractability, the study of the performance of mobile wireless networks has considered the cell dwell time and the channel holding time to be negative exponentially distributed RVs. Nevertheless, recent papers [35], [49] have concluded that to capture the overall effects of cellular shape and user’s mobility patterns, both the channel holding time and the cell dwell time need to be modeled as RVs with more general distribution. In addition, the large degree of variability in size and shape of microcells suggests that a more general model for the cell dwell time is needed [15], [35]–[51]. In [37], it has been shown that the channel holding time is not exponentially distributed. In fact, there are empirical studies showing that the channel holding times are gamma or lognormal distributed. For instance, in [42], Zonoozi and Dassanayake show that for a generalized set of mobility conditions, the cell dwell time and the channel holding time are well characterized by a gamma distribution. In [15], [50], and [51], it has been shown by empirical studies that the channel holding times are environment dependent, but all types of channel occupancy times can very closely be approximated by a lognormal distribution. Furthermore, the authors of [40], [45], and [46] stated that the time variables that describe the users’ mobility in wireless communications networks are best modeled by using phase-type distributions. To make the mathematical analysis more tractable when generally distributed time variables are considered, the authors of [46]–[49] proposed the use of a Laplace transform as a mathematical tool to analyze the system performance. Moreover, closed expressions for many performance metrics can be derived when the Laplace transform of the probability density function (pdf) of the channel holding time is a rational function. The Laplace transforms of the phase-type probability distributions are examples of rational functions. Consequently, these types of distributions allow a very simple mathematical analysis, and more importantly, one of them (i.e., Mixed-Erlang) has the universal approximation capability, that is, it is able to arbitrarily closely approximate to the distribution of any nonnegative RV as well as measured data [13]. Thus, general and easily computable results can be obtained if the involved time variables are assumed to be phase-type distributed. In this paper, a mathematical analysis for obtaining the system-level performance metrics is developed considering that the cell dwell time and the unencumbered call-interruption time are phase-type distributed RVs.

As stated in the previous sections, in this paper, the effect of link unreliability is captured through the appropriate characterization of the probability distribution of the unencumbered call-interruption time. This time should adequately be characterized in a statistical way through its pdf. The pdf of the unencumbered call-interruption time due to link unreliability can be obtained by two different ways, depending on the level of the statistics available. In this paper, both link- and system-level approaches are addressed. Link-level statistics can be obtained from a specific wireless channel model (i.e., Gilbert–Elliot [52], [53], Fritchman [54], etc.), and system-level statistics (i.e., key performance indicators (KPIs) and counters) can easily be obtained at BSs in real cellular networks [15]. Referring to the link-level approach, in Section V, the pdf of the unencumbered call-interruption time is obtained considering the Fritchman channel model. When the system-level approach is considered, a functional relationship between the call-interruption processes of the analytical model and some system-level
statistics is derived (this is performed in Section VI). The main conclusion from Sections V and VI is that it is imperative to model the unencumbered call-interruption time by a general probability distribution. It is important to notice that the mathematical analysis developed here is based on the analytical framework proposed by Fang in [46]–[48] for the performance evaluation of wireless networks and mobile computing systems. Nevertheless, in Fang’s work, the important issue of link unreliability is ignored, that is, in Fang’s work, call forced-termination is considered to occur due only to resource insufficiency.

Therefore, there is a need for a new reformulation of teletraffic analysis in cellular systems to take into account link unreliability under more realistic considerations. This is the aim of this paper. In particular, our contributions can be summarized as follows. Teletraffic analysis is performed by taking into account both resource insufficiency and link unreliability at the system level. More importantly, most of the involved times are considered generally distributed RVs. The effect of link unreliability is captured through a system-level model by introducing a simple call-interruption process. This process has an associated potential generally distributed RV that is called “unencumbered call-interruption time.” Thus, the effect of link unreliability is captured through the appropriate probabilistic characterization of the unencumbered call-interruption time. In addition, of practical importance, it is shown how the pdf of this time variable can be derived in terms of easily measurable system-level parameters of real cellular networks such as KPIs and counters [15]. Additionally, we also show how the pdf of the unencumbered call-interruption time can be derived in terms of link-level statistics (considering either the Gilbert–Elliott [52], [53] or Fritchman [54] channel model).

III. PROPOSED PROCESS TO MODEL THE CALL INTERRUPTION DUE TO LINK UNRELIABILITY

In this section, the proposed model that includes call dropping due to link unreliability in the performance evaluation of mobile cellular systems is presented. First, the mathematical notation and the used time variables are explained and defined, respectively. Unless otherwise stated, the RVs are represented with bold uppercase letters (i.e., \( \mathbf{X} \)). The expected value of \( \mathbf{X} \) is denoted with \( E[\mathbf{X}] \), and the scalars are denoted with cursive letters.

First, let us define the unencumbered service time per call \( x_s \) (also known in the literature as the call holding time [4], [31]–[33]) as the duration of a requested call connection, and it is equivalent to the call duration in the fixed telephone network. It has widely been accepted in the literature that the unencumbered service time for voice calls can adequately be modeled by a negative exponentially distributed RV [1]–[5], [31]–[34], [40], [44], [46]. The RV used to represent this time is \( \mathbf{X}_s \).

Now, the cell dwell time or the cell residence time \( x_d^{(j)} \) is defined as the time that an MS spends in the \( j \)th (for \( j = 0, 1, \ldots \)) handed off cell, irrespective of whether it is engaged in a call (or session). The RVs used to represent these times are \( \mathbf{X}_d^{(j)} \) (for \( j = 0, 1, \ldots \)) and are assumed to be independent and identically generally distributed (IID). 3 For homogeneous cellular systems, this assumption has widely been accepted in the literature [31]–[48].

The residual cell dwell time \( x_r \) is defined as the time between the instant that a new call is initiated and the instant that the user is handed off to another cell. Notice that the residual cell dwell time is only defined for new calls. The RV used to represent this time is \( \mathbf{X}_r \). Thus, the pdf of the residual time \( f_{\mathbf{X}_r(t)} \) can be calculated in terms of the cell dwell time using the residual life theorem [55]

\[
f_{\mathbf{X}_r(t)} = \frac{1}{E[\mathbf{X}_d]} \left[ 1 - F_{\mathbf{X}_d}(t) \right]
\]

where \( E[\mathbf{X}_d] \) is the mean value of the RV \( \mathbf{X}_d \), and \( F_{\mathbf{X}_d}(t) \) is the cumulative probability distribution function (cdf) of the cell dwell time.

The channel holding time \( x_c^{(j)} \) is the time that a call occupies a given set of radio resources in the \( j \)th (for \( j = 0, 1, \ldots \)) handed off cell before its call is either completed, handed off to another cell, or interrupted due to link unreliability. Notice that the channel holding time \( x_c^{(j)} \) could easily be measured at the BS in real cellular networks. In fact, the total channel holding time (i.e., the aggregated channel holding time \( x_c^{(j)} \) in the different cells a user roams until its call is either concluded or interrupted due to either resource insufficiency or link unreliability) is used for billing purposes.

However, notice that in mobile networks calls are not always successfully completed.

Based on the two different physical processes in which an interrupted call is involved (resource insufficiency and link unreliability), the call-interruption process is explained as follows.

1) After a new call is originated or a handoff is requested and accepted by a BS, the physical link between BS and MS may become degraded during the call and forced to terminate due to link unreliability. The period of time between the instant that the call originated and the moment that the call is forced to terminate due to link unreliability is called the channel holding time for calls forced to terminate due to link unreliability, and it is denoted by \( x_{cd} \). The RV used to represent this time variable is \( \mathbf{X}_{cd} \).

2) On the other hand, when a call attempts to handoff to another cell, there could be insufficient resources in the target cell and thus would be forced to terminate due to resource insufficiency. The period of time between the instant that the call originated in a specific cell and the moment that the call is forced to terminate due to resource insufficiency is called the channel holding time for calls forced to terminate due to resource insufficiency, and it is denoted by \( x_{ch} \). The RV used to represent this time is \( \mathbf{X}_{ch} \).

3) To simplify the notation, the super index \( j \) indicates that the referred cell has been omitted because these variables are considered to be IID. A similar rule is applied for variables \( \mathbf{X}^{(j)}_d \) and \( \mathbf{X}^{(j)}_r \).

4) Additionally, as noticed in [15], by using the Clear Codes reported in the databases of the network operator, calls are classified in dropped and not dropped, distinguishing for the former the causes of dropping.
The teletraffic model proposed in this paper considers an interruption process and a potential associated time to this process, which is called “unencumbered call-interruption time” to take into account the link unreliability. The unencumbered call-interruption time is denoted by $x^{(j)}$, and it is defined as the period of time from the epoch that the MS establishes a link with the $j$th handed-off cell (for $j = 0, 1, 2, \ldots$) until the instant that the call would be interrupted due to link unreliability assuming that the MS has neither successfully completed his call nor been handed off to another cell. Physically, this time represents the period in which a call would be terminated under the assumption that both the cell dwell time and the unencumbered service time are of infinite duration. The RVs used to represent these times are $X^{(j)}_j$ (for $j = 0, 1, 2, \ldots$) and are assumed to be IID and generally distributed.

The call-interruption time is said to be “unencumbered” because the interruption of a call in progress by link unreliability can or cannot occur, depending on the values of the cell dwell time and the unencumbered service time. Thus, the unencumbered call-interruption time only depends on the link reliability. Notice that the unencumbered call-interruption time is a “potential” time, that is, it could not directly be measured in real cellular networks because it only depends on the wireless channel conditions. The next section shows how the model for taking into account the link unreliability is incorporated in the teletraffic analysis. Then, mathematical expressions for several system-level performance metrics are derived.

IV. SYSTEM MODEL

A. Teletraffic Model

In this section, the general guidelines for mathematical analysis are given. A homogeneous multicellular system with omnidirectional antennas located at the center of each cell is assumed. Each cell has a maximum number of radio channels $C$. As it has widely been accepted, a new call arrival process is assumed to follow a Poisson process with a mean arrival rate $\lambda_n$ per cell.

In the evaluation environment, it is assumed that each cell receives handoff calls from six different cells. The handoff call arrival process generated by a single cell is clearly not Poisson. However, the combined process from six different neighboring cells can adequately be approximated by a Poisson process [56]–[58]. To further support the assumption of a Poisson process, it was demonstrated in [58] that considering a smooth process instead of a Poisson process has an insignificant impact on the call blocking probabilities (i.e., new call blocking, call forced-termination and handoff failure probabilities) when these probabilities are around the nominal load (i.e., about 2%–5%, which are the required values in a wireless network design).

It is important to mention that the teletraffic analysis developed here is applicable to work being done in many areas of backbone networks [Multiprotocol Label Switching (MPLS), Asynchronous Transfer Mode (ATM), Transmission Control Protocol/Internet Protocol (TCP/IP)]. As in [21], the access network is assumed to use connection-oriented resource allocation to provide varying levels of QoS. Thus, in this paper, it is assumed that connection-oriented mechanisms are used in access networks to limit the number of connections that inject traffic into backbone networks.

B. Channel Model

To characterize the unencumbered call-interruption time by means of link-level statistics (in terms of its pdf and mean), two discrete channel models are used, that is, Fritchman and Gilbert–Elliott. These models have widely been used to capture the periods of signal degradation [9]–[14], [27]. In the Gilbert–Elliott model (two-state Markov chain model), the state space of the wireless channel consists of $\Omega = \{good, bad\}$. The transition probabilities are defined as follows. Given that the current state is a good one, the probability that the next state is a good one is denoted by $(p_G)$, and the probability that the next state is a bad one is denoted by $(1 - p_G)$. On the other hand, given that the current state is a bad one, the probability that the next state is a bad one is denoted by $(p_B)$, and the probability that the next state is a good one is denoted by $(1 - p_B)$. Then, the channel is modeled by a sequence of alternating good and bad states. For this model, a channel cycle is defined as the continuous good state and its next consecutive bad state. The good time $x^{(k)}_g$ and bad time $x^{(k)}_b$ variables are defined as the time duration of the good and bad states of the $k$th cycle, respectively. The RVs used to represent these times are $X^{(g)}_k$ and $X^{(b)}_k$ (for $k = 1, 2, \ldots$), respectively. Let us consider that state transitions are made after every channel symbol, and that the symbol duration is fixed. Notice that the sojourn time in the good (bad) state takes the value of $n$ times symbol duration with probability $p_G^n (1 - p_G) [(1 - p_B) (1 - p_B)]$. Consequently, the state sojourn time is geometrically distributed. However, for a high probability of staying within a state and a large number of channel symbols, the state sojourn time could be approximated by a negative exponentially distributed RV [59].

On the other hand, the Fritchman channel model [54] is a generalized Gilbert–Elliott model in which the channel is modeled by a sequence of several good or bad states. The state sojourn time is determined by the transition probabilities. Then, the time that the channel stays in each state can be assumed to be independent and generally distributed. In some cases, a continuous sequence of good (bad) states could be represented as only one state with arbitrary distributed time duration. In this paper, to connect the derived expressions to some practical use,
three different probability distributions are considered for the time that the link spends in each state, that is, hyperexponential, \(n\)-stage Erlang, and negative exponential.

As in [9]–[11], for these two channel models, we assumed that the calls in the good channel state can communicate with error free and those in the bad channel state may fail, but not definitely, to complete the conversation or session owing to the broken link. When the channel turns from good state to bad state, the call may not immediately be dropped because of the link layer error protection scheme. In many wireless cellular standards (i.e., Global System for Mobile Communications (GSM) and UMTS), there is usually a timer, named Monitoring Channel Timer (MCT), with duration \(T_{mc}\), in the link layer to monitor the channel state; this time variable is considered to be an exponentially distributed RV [15]. The RV used to represent this time is \(T_{mc}\). In this paper, we take into account the link reestablishing procedure described in [11] and [12].

The authors of [11] and [12] state that as the channel becomes worse and no messages are exchanged between MS and BS during the duration \(T_{mc}\), a link reestablishing procedure is performed. This procedure works as follows. When the BS detects the link broken state, a number of consecutive timers with length \(T_{LER}\) are started to try relinking with the MS. The link reestablishing procedure is performed a maximum of \(\text{NLER}\) times. If the duration of the bad state is greater than \(T_{mc}\) plus \(T_{LER} \times \text{NLER}\), it is considered that the link is not reliable.

V. Determination of the Probability Distribution of the Unencumbered Call- Interruption Time

From a Link-Level Channel Model

In this section, the pdf of the unencumbered call-interruption time \(X^{(i)}\) is found by considering the Fritchman channel model. Equations (3)–(10) are original expressions not previously reported in the literature.

To analytically find the pdf of the unencumbered call-interruption time, it is necessary to define \(P_{r}\) as the probability that the link could not be reestablished during the bad state. This probability is mathematically expressed as [11]

\[
P_r = P\{X_b > T_{LER} \times \text{NLER} + T_{mc}\}. \tag{2}
\]

The unencumbered call-interruption time is the aggregated time that a call spends in the good and bad channel states. Notice that a call should always start in a good state because a call could not be established if the channel is in a bad state. On the other hand, if a call is not interrupted in the bad state of a certain cycle, it should go through the good state of the next cycle before it could be interrupted. By using the good time \(x_g^{(k)}\) and bad time \(x_b^{(k)}\) variables defined in Section IV-B, the unencumbered call-interruption time \((x_i)^{8}\) for calls interrupted in the first cycle is given by \(x_g^{(1)} + x_b^{(1)}\), which occurs with probability \(P_r\) (if the reestablishing procedure is unsuccessful). If the reestablishing procedure is successful in the first cycle and unsuccessful in the second cycle, the unencumbered call-interruption time is given by \(x_g^{(2)} + x_b^{(2)}\) with probability \((1 - P_r)P_r\), and so on. Thus, the pdf of the unencumbered call-interruption time can be expressed in terms of the pdf of the good and bad states as

\[
f_{X_i}(t) = f_{X_g^{(i)} + X_b^{(i)}}(t)P_r + f_{X_g^{(i)} + X_b^{(i)} + X_g^{(2)} + X_b^{(2)}}(t)(1 - P_r)P_r + \cdots + \sum_{k=1}^{n} f_{X_g^{(k)} + X_b^{(k)}}(t)(1 - P_r)^{n-1}P_r + \cdots \tag{3}
\]

where \(f_{X_g^{(k)} + X_b^{(k)}}(t)\) denotes the pdf of the summation of \(n\) IID RVs, each of them representing the aggregated time spent in the good and bad states in the \(k\)th channel cycle. As communication systems are designed to be reliable, the mean time that a channel is in the bad state is usually much less than the mean time that the channel is in the good state (i.e., typically \(E[X_g^{(k)}] \gg E[X_b^{(k)}]\)). Consequently, the pdf of the unencumbered call-interruption time can be approximated by

\[
f_{X_i}(t) \approx f_{X_g^{(1)}(t)}P_r + f_{X_g^{(1)} + X_b^{(2)}(t)}(1 - P_r)P_r + \cdots + \sum_{k=1}^{n} f_{X_g^{(k)}(t)}(1 - P_r)^{n-1}P_r + \cdots \tag{4}
\]

where \(f_{\sum_{k=1}^{n} X_g^{(k)}(t)}\) denotes the pdf of the summation of \(n\) IID RVs, each of them representing the time spent in the good state in the \(k\)th channel cycle. Since the RVs used to represent the time spent in a good state for each cycle are assumed to be IID (i.e., \(X_g = X_g^{(1)} = X_g^{(2)} = \cdots = X_g^{(k)}\)), expressing (4) can be rewritten in a compact form as

\[
f_{X_i}(t) \approx \sum_{n=1}^{\infty} \sum_{k=1}^{n} f_{X_g^{(k)}}(t)(1 - P_r)^{n-1}P_r. \tag{5}
\]

Since the RVs \(X_g\) are independent, the pdf of their sum can be obtained as the convolution of their pdfs [55]. Thus, it is possible to use Laplace transform to obtain a closed expression for the pdf of \(X_i\) [46]–[49], that is

\[
f_{X_i}(s) \approx \sum_{n=1}^{\infty} \left( f_{X_g^{(1)}}(s) \right)^n (1 - P_r)^{n-1}P_r, \tag{6}
\]

where \(f_{X_g^{(1)}}\) represents the Laplace–Stieljes transform

This summation converges to

\[
f_{X_i}(s) \approx \frac{f_{X_g^{(1)}}(s)P_r}{1 - f_{X_g^{(1)}}(s)(1 - P_r)}. \tag{7}
\]

When \(X_g\) is considered negative exponentially distributed (i.e., when the channel is modeled by a Gilbert–Elliot channel

Notice that the Gilbert–Elliot channel model is a special case for the Fritchman channel model in which the time that the channel is in either bad state or good state is modeled by a negative exponentially random variable.

The super index \((j)\) was omitted in \(X_i\) since it is considered that the unencumbered call-interruption time does not depend on the visited cell.
model) with parameter $\mu_g$, it is easy to show that the distribution for $X_i$ is given by

$$f_X^*(s) \approx \frac{\mu_g}{s + \mu_g} P_r = \frac{\mu_g P_r}{s + \mu_g P_r}. \quad (8)$$

Then, in this case, $X_i$ approximately follows a negative exponential distribution with mean $1/(\mu_g P_r)$.

Erlang and hyperexponential distributions are also used for the state sojourn time as examples to find the characterization of the unencumbered call-interruption time $X_i$ by using the Fritchman model. From (7), it is easy to obtain a closed form for the Laplace transform of the pdf of the unencumbered call-interruption time by considering that the good state duration has an $n$-order Erlang or a hyperexponential distribution. However, this transform depends on the order $n$ of the Erlang (hyperexponential) distribution, and therefore, the inverse Laplace transform will also depend on this parameter. Equations (9) and (10) show Laplace transforms of the pdf of the unencumbered call-interruption time considering that the good state duration has an $n$-order Erlang or hyperexponential distribution

$$f_X^*(s) \approx \frac{\mu^n}{(s + \mu)^n} P_r = \frac{\mu^n P_r}{s + \mu^n (1 - P_r)} \quad (9)$$

where $\mu$ represents the parameter of each stage of the Erlang distribution

$$f_X^*(s) \approx \sum_{l=1}^{m} \frac{\alpha_l \mu^l}{(s + \mu^l)} P_r \quad (10)$$

where $\mu_l$ represents the parameter of the $l$th phase, $\alpha_l$ represents the probability of choosing the $l$th phase, and $m$ represents the number of phases of the hyperexponential distribution. From (9) and (10), it is evident that the inverse transform will depend on the order of the distribution of the good state duration.

To validate analytical results and evaluate the accuracy of the approximation, numerical and simulation results are provided in this section. For numerical evaluation, the same values of the parameters used in [9]–[11] were employed; that is, $N_{LER} = 3$, $TLER = 1000$ ms,$^9$ mean value of $T_{mnc} = 6$ s, $E[X_g] = 24$ s, and $E[X_b] = 3$ s.$^{10}$ Fig. 1 shows the pdf of the unencumbered call-interruption time $X_i$ obtained via simulation and using the approximation given by (8). Different distributions for good and bad channel state duration were considered in this figure. Table I shows the mean and coefficient of variation (CV)$^{11}$ for the good- and bad-state channel duration, and the approximated unencumbered call-interruption time. In Fig. 1 and Table I, the hyperexponential, Erlang, and negative exponential probability distributions of the channel state duration are labeled H, E, and N, respectively; each one of them is followed by its corresponding value of the CV. In Fig. 1, the simulation and analytical results are labeled with “S” and “A,” respectively. From Fig. 1, it is observed that the analytical and simulation results match very well, which validates our approximations and analytical model. In addition, from Fig. 1, it is important to observe that the pdf of the unencumbered call-interruption time strongly depends on the distribution of the good state duration.

On the other hand, from Fig. 1 and Table I, it is observed that the pdf of the unencumbered call-interruption time follows many different forms. Consequently, we conclude that, to obtain more realistic results, it is more reasonable to study the system-level performance assuming that the unencumbered call-interruption time is modeled as a generally distributed RV.

From Table I, it is observed that when the state sojourn time is considered to be hyperexponentially distributed, the CV that has a greater impact on the mean value of the unencumbered call-interruption time is that of the bad state duration (i.e., a change in the CV of the distribution of the “good time” does

9Practically, $TLER$ is very small since its value is normally determined under a good propagation environment (i.e., small losses).

10Notice that for accurate system-level characterization of the unencumbered call-interruption time, it is important to adequately characterize the Fritchman channel model from link-level performance evaluation.

11The CV is defined as the ratio between the standard deviation and the mean of a random variable, and this parameter is an index of data dispersion around the mean value. For negative exponentially, hyperexponentially, and Erlang-distributed random variables, the CV is $CV = 1$, $CV \geq 1$, and $CV \leq 1$, respectively.
Fig. 2. Call scenario when the call is successfully completed.

Fig. 3. Call scenario when the call is forced to terminate due to resource insufficiency.

not have a great impact in the mean of the unencumbered call-interruption time), whereas the state sojourn time distributed as \( n \)-Erlang does not have an impact on the mean value of the unencumbered call-interruption time. In contrast to this, the CV of the distribution of the good state sojourn time has a greater impact on the CV of the distribution of the unencumbered call-interruption time. Notice that when the distribution of the bad state duration is hyperexponential, the CV of the unencumbered call-interruption time is greater than 1 (i.e., it can adequately be modeled by a hyperexponential distribution). On the other hand, when the bad state duration has an Erlang distribution, the CV is less than 1 (i.e., it can adequately be modeled by an Erlang distribution).

VI. DETERMINATION OF THE PROBABILITY DISTRIBUTION OF THE UNENCUMBERED CALL-INTERRUPTION TIME FROM KPIS AND COUNTERS

This section shows how the pdf of the unencumbered call-interruption time \( X^{(j)} \) can be obtained from the channel holding time probability distribution. As stated before, the channel holding time can easily be measured at BSs in real mobile wireless networks from KPIS.

To model the channel holding time, it is necessary to express it as a function of the unencumbered service, cell dwell, unencumbered call-interruption, and residual cell dwell times. Then, when a call is in progress in the \( j \)th handed-off cell, there are three different possible realizations for the channel holding time: 1) It takes the value of the service time; 2) it takes the value of the unencumbered call-interruption time; or 3) it takes the value of the cell dwell time. As a consequence, it is necessary to distinguish between the channel holding time for successfully terminated calls, forced to terminate calls due to link unreliability, and forced to terminate calls due to resource insufficiency (handed off calls). These three different call scenarios and the relationship among the involved time variables are graphically represented by the time diagrams shown in Figs. 2–4, respectively. Please observe that the different time variables used in Figs. 2–4 are defined in Section III. To differentiate these three situations, let us describe each in detail.

1) Fig. 2 shows the time diagram for the case when a call is successfully completed in the \( j \)th handed-off cell. In this situation, the service time is smaller than both the cell dwell time and the unencumbered call-interruption time in the \( j \)th handed-off cell. Thus, because of the assumption that the unencumbered service time has a memoryless distribution, the channel holding time in the \( j \)th handed-off cell takes the value of its corresponding service time. That is, \( x^{(j)}_c = x_s \) if and only if (iif) \( x_s < \min(x^{(j)}_s, x^{(j)}_d) \) for \( j > 0 \), or \( x^{(0)}_c = x_s \) if \( x_s < \min(x^{(0)}_s, x_r) \).

2) Fig. 3 shows the time diagram for the case when a call is forced to terminate due to link unreliability in the \( j \)th handed-off cell. In this situation, the unencumbered call-interruption time is smaller than both the cell dwell time and the unencumbered call-interruption time in the \( j \)th handed-off cell.
and the service time in the jth handed-off cell. Thus, the channel holding time in the jth cell takes the value of its corresponding unencumbered call-interruption time.

That is, $x_c^{(j)} = x_d^{(j)}$ if $x_d^{(j)} < \min(x_s, x_c^{(j)})$ for $j > 0$, or $x_c^{(0)} = x_r^{(0)}$ if $x_r^{(0)} < \min(x_s, x_r)$.

3) Finally, Fig. 4 shows the time diagram for the case when a call in the jth handed-off cell is handed off to another cell. In this situation, the cell dwell time is smaller than both the unencumbered call-interruption time and the service time in the jth handed-off cell. Thus, the channel holding time in the jth handed off cell takes the value of its corresponding cell dwell time. That is, $x_c^{(j)} = x_d^{(j)}$ if $x_d^{(j)} < \min(x_s, x_c^{(j)})$ for $j > 0$, or $x_c^{(0)} = x_r$ if $x_r < \min(x_s, x_r^{(0)})$.

Thus, the RV $X_c^{(j)}$, which models the channel holding time in the jth handed-off cell, is defined as

$$X_c^{(j)} = \begin{cases} \min\left(X_s^{(0)}, X_s, X_r\right) ; & j = 0 \\ \min\left(X_c^{(j)}, X_s, X_d^{(j)}\right) ; & j > 0. \end{cases}$$

(11)

Notice that the interruption process in the proposed mathematical model should statistically be characterized by the pdf of its associated unencumbered call-interruption time. Unfortunately, this time cannot directly be determined from real cellular networks because it only represents a hypothetical event. That is, considering a call in progress, if both call’s unencumbered service time and cell dwell time were considered infinite, the call in progress will be dropped at the unencumbered call-interruption time (due to wireless link unreliability).

For practical usefulness of our proposed analytical model, a functional relationship between the unencumbered call-interruption time $X_c^{(j)}$ and the system-level statistics (i.e., KPIs and counters) is developed. In particular, the statistics of the channel holding time for calls forced to terminate due to link unreliability (which can be measured in a real cellular network) can be represented by

$$F_{X_{sd}}(t) = P\{X_i \leq t|X_i \leq \min(X_s, X_r)\}. \tag{12}$$

Then, to characterize the interruption process of the proposed mathematical model, it is necessary to relate the probability distribution of the channel holding time for calls forced to terminate due to link unreliability [i.e., (12)] with the probability distribution of the unencumbered call-interruption time (i.e., $P\{X_s \leq t\}$). Notice that it is necessary to differentiate between new and handed off calls because of the different probability distributions of cell dwell time and residual cell dwell time. When new calls are analyzed, the residual cell dwell time should be considered, whereas the cell dwell time should be considered when the handed-off calls are analyzed. First, let us characterize the unencumbered call-interruption time of new calls by using Bayes’ theorem as

$$P\{X_i \leq t\} = \frac{P\{X_i \leq t|X_i \leq \min(X_s, X_r)\}P\{X_i \leq \min(X_s, X_r)\}}{P\{X_i \leq \min(X_s, X_r)\}|X_i \leq t}. \tag{13}$$

The argument of the first factor in the numerator on the right hand of (13) represents the cdf of the channel-holding time for calls forced to terminate due to link unreliability, and the second factor in the numerator on the right hand of (13) represents the probability that calls are forced to terminate due to link unreliability. Both could easily be obtained from system-level measurements. The denominator on the right hand of (13) represents the probability that calls are forced to terminate due to link unreliability conditioned on the unencumbered call-interruption time. This term depends on the probability distribution of the unencumbered call-interruption time, which is precisely the distribution that should be determined.

To approximate the required distribution by a general non-negative probability distribution in an iterative way, the following general algorithm is proposed.

Step 0) From system-level measurements and using a probabilistic fitting procedure (i.e., the moment matching technique [60], [61]), the cdf of the channel-holding time for calls forced to terminate due to the link unreliability given by (12) can be found.
Step 1) Propose a general probability distribution to characterize the unencumbered call-interruption time \( X^{(j)}_i \).
Step 2) Choose a set of values for the parameters of the proposed probability distribution of \( X^{(j)}_i \).
Step 3) Prove the accuracy of this distribution.

To prove the accuracy of the suggested approximation, the “goodness of fit” for this approximation is defined as

\[
G = \frac{\int_0^{\infty} \left| \frac{\hat{P}\{X_i \leq t\} - \frac{F_{X^{(j)}_i}}{\hat{F}_{X^{(j)}_i}}}{\hat{F}_{X^{(j)}_i}} \right|^2 dt}{2 \int_0^{\infty} \left| \hat{P}\{X_i \leq t\} \right|^2 dt}
\] (14)

where \( G \) indicates the normalized square error between two functions, which is on the interval \([0, 1]\). Then, \( G = 0 \) specifies an exact fit, and \( G = 1 \) indicates no fit.\(^\text{13}\)

If \( G > \delta \) (where \( \delta \) is the proposed accuracy for the approximation), go to Step 2. Otherwise, the proposed accuracy is achieved, and the process ends. A similar procedure can be followed when the handoff off calls are analyzed. In this case, the cell dwell time should be considered instead of the residual cell dwell time. The complexity of this algorithm depends on the particular distributions and the number of parameters.

Although the complexity of this algorithm depends on the particular distribution used and the number of its parameters, in general, this algorithm has a high computational cost. This is because it could be necessary to run the proposed algorithm with several distributions to find the one that best fits. Then, efforts need to be made to provide some guidelines in how to choose a set of values for the parameters of the proposed pdf of \( X_i \) in Step 2. This guideline is important to guarantee that this algorithm converges in a reasonable time. However, optimization and improving the computational cost of this algorithm are out of the scope of this paper and left as material for future research.

### VII. Mathematical Analysis for System-Level Performance Metrics

In this section, the proposed mathematical model to include call dropping due to link unreliability is used to evaluate the performance of mobile cellular networks by considering that the unencumbered call-interruption time and the cell dwell time are general distributed RVs. Mathematical expressions for several system-level performance metrics are derived (i.e., handoff probability, interruption probability due to link unreliability, handoff rate, and call forced-termination probability).

#### A. Minimum Time Variable Transformations

Let us define the pdf of the minimum of some pairs of RVs that will be used in the rest of this paper. Consider the case when calls are forced to terminate due to link unreliability [i.e., \( X^{(j)}_i < \min(X_s, X^{(j)}_d) \)]. Let us define the RV \( Y^{(j)} \) as the minimum of the service time and the unencumbered call-interruption time in the \( j \)th handed off cell

\[
Y^{(j)} = \begin{cases} \min(X_s, X^{(j)}_d); & j = 0 \\ \min(X_s, X^{(j)}_d); & j > 0. \end{cases}
\] (15)

The pdfs of these RVs are given by [55]

\[
f_{Y^{(j)}(t)} = \begin{cases} f_{X^{(j)}_s}(t) + f_{X^{(j)}_d}(t) - f_{X^{(j)}_s}(t)f_{X^{(j)}_d}(t); & j = 0 \\ -f_{X^{(j)}_s}(t)f_{X^{(j)}_d}(t); & j > 0. \end{cases}
\] (16)

Finally, consider the case when calls are forced to terminate due to handoff failure [i.e., \( X^{(j)}_i < \min(X_s, X^{(j)}_d) \)]. Let us define the RV \( Z^{(j)} \) as the minimum of the service time and the unencumbered call-interruption time in the \( j \)th handed off cell

\[
Z^{(j)} = \min\left(X_s, X^{(j)}_d\right); \quad j \geq 0.
\] (17)

The pdfs of these RV are given by [53]

\[
f_{Z^{(j)}(t)} = f_{X^{(j)}_s}(t) + f_{X^{(j)}_d}(t) - f_{X^{(j)}_s}(t)f_{X^{(j)}_d}(t) - f_{X^{(j)}_s}(t)F_{X^{(j)}_d}(t); \quad j \geq 0.
\] (18)

#### B. Handoff Probability

The handoff probability (\( P_H \)) is defined as the probability that a call needs to be handed off during its remaining service time. This quantity is very important to evaluate the performance of the system since a call could be forced to terminate due to resource insufficiency in a handoff attempt.

A call is handed off to another cell when the cell dwell time is smaller than the unencumbered service time and the unencumbered call-interruption time in the \( j \)th handed-off cell. Considering both new calls and handed-off calls and the RV \( Z \) defined in the previous section, the mathematical expression for the handoff probability is given by

\[
P^{(j)}_H = \begin{cases} P\{X_i < Z^{(j)}\}; & j = 0 \\ P\{X^{(j)}_d < Z^{(j)}\}; & j > 0. \end{cases}
\] (19)

By using the moment generator function of the pdf of the involved RVs and a similar procedure to that described in [46]–[49], the handoff probability can be expressed by

\[
P^{(j)}_H = \begin{cases} \frac{1}{2\pi\sigma^j} \int_{-\infty}^{+\infty} f_{X^{(j)}_s}(s) f_Z^{(j)}(-s) ds; & j = 0 \\ \frac{1}{2\pi\sigma^j} \int_{-\infty}^{+\infty} f_{X^{(j)}_d}(s) f_Z^{(j)}(-s) ds; & j > 0. \end{cases}
\] (20)

where \( f_{X^{(j)}_s}(s) \) represents the Laplace–Stieljies transform\(^\text{14}\) of the pdf of \( X_i \). When \( f_{X^{(j)}_s}(s) \), \( f_{X^{(j)}_d}(s) \), and \( f_Z^{(j)}(s) \) are rational

\(^\text{13}\)Other approximation criteria could be used (i.e., minimizing Kolmogorov–Smirnoff criterion, chi-square, etc.) [40].

\(^\text{14}\)The Laplace–Stieljies transform of a pdf is the moment generator function evaluated in \(-s\).
functions in $s$, the well-known residue theorem can be applied to find $P_H^{(j)}$ as follows [46]:

$$P_H^{(j)} = \begin{cases} - \sum_{p \in \sigma_p} \text{Res}_{s=p} \frac{f_{X^{(j)}_s}(s)}{s} f_{Z}(-s); & j = 0 \\ - \sum_{p \in \sigma_p} \text{Res}_{s=p} \frac{f_{X^{(j)}_s}(s)}{s} f_{Z}(-s); & j > 0 \end{cases}$$  

(21)

where $\sigma_p$ is the set of poles of $f_{Z}(-s)$, and $\text{Res}_{s=p}$ represents the residue at pole $s = p$. Considering that $X^{(j)}_d$ (for $j = 1, 2, \ldots$) are identically distributed RVs, then $P_H^{(j)} = P_H^{(i)} = P_H$ for $j \neq i$, and $P_H^{(0)}$ depends on the residual cell dwell time.

C. Interruption Probability Due to Link Unreliability

As previously stated, the other fundamental cause of call forced-termination is link unreliability. Then, to adequately evaluate the system performance, it is necessary to obtain the call-interruption probability. Considering the proposed model, a call will be forced to terminate due to link unreliability when the unencumbered service time and the cell dwell time in the $j$th handed-off cell. Then, considering the RV $Y^{(j)}$ defined in Section VII-A, mathematically, the probability that a call would be forced to terminate due to link unreliability can be expressed as

$$P^{(j)} = P\left\{ X^{(j)}_i < Y^{(j)} \right\}; \quad j = 0, 1, 2, \ldots.$$  

(22)

By using the moment generator function of the pdf of the RVs involved and a similar procedure to those described in [46]–[49], the handoff probability could be calculated by

$$P^{(j)} = \frac{1}{2\pi j} \int_{-\infty}^{\infty} f_{Z}^{*}(s) \frac{f_{X^{(j)}_s}(s)}{s} f_{Y^{(j)}(-s)} ds; \quad j = 0, 1, 2, \ldots.$$  

(23)

When $f_{X^{(j)}_s}(s)$ and $f_{Y^{(j)}(-s)}$ are rational functions in $s$, then the well-known residue theorem can be applied to easily find $P^{(j)}$ as

$$P^{(j)} = - \sum_{p \in \sigma_p} \text{Res}_{s=p} \frac{f_{X^{(j)}_s}(s)}{s} f_{Y^{(j)}(-s)}; \quad j = 0, 1, 2, \ldots.$$  

(24)

where $\sigma_p$ is the set of poles of $f_{Y^{(j)}(-s)}$. Considering that $X^{(j)}_d$ (for $j = 0, 1, 2, \ldots$) are identically distributed RVs, then $P^{(j)} = P^{(i)} = P_I$ for $j \neq i$, and $P_I^{(0)}$ depends on the residual cell dwell time.

D. Handoff Rate

The handoff rate ($\lambda_h$) is defined as the average number of handoffs undertaken during the actual call connection in a wireless cellular network. Thus, mathematically, it can be expressed as

$$\lambda_h = E\{H\} \lambda_n.$$  

(25)

The mean value of handoffs could be obtained as follows:

$$E\{H\} = \sum_{k=0}^{\infty} k P\{H = k\}$$  

(26)

where $P\{H = k\}$ is the probability that a call suffers $k$ handoffs. Mathematically, it could be expressed in terms of the channel holding time and the handoff probability $P_H$, which is the same in all cells with $j > 0$. It could be noticed that $H = 0$ if the unencumbered service time or the unencumbered call-interruption time is shorter than the residual cell dwell time (i.e., with probability $1 - P_H^{(0)}$). That is, the call is successfully terminated or forced to terminate due to link unreliability before the mobile moves out of the initial cell. On the other hand, an admitted call experiences $k$ handoffs during its call connection life ($H = k$) if the call is either successfully completed having previously experienced $k - 1$ handoffs ($k > 0$) or due to link unreliability in the $(k + 1)$th visited cell or successfully having $k - 1$ handoffs ($k > 0$) and then failed in the $k$th handoff attempt. If the handoff failure probability is represented by $P_H$, then the probability distribution of the number of experienced handoffs can be written as

$$P\{H = 0\} = 1 - P_H^{(0)}$$

$$P\{H = 1\} = P_H^{(0)} \left[ P_H + (1 - P_H)(1 - P_H) \right]$$

$$P\{H = 2\} = P_H^{(0)} (1 - P_H) P_H^{(1)} \left[ P_H + (1 - P_H)(1 - P_H) \right]$$

$$\vdots$$

$$P\{H = k\} = P_H^{(0)} P_H^{k-1}(1-P_H) \left[ P_H + (1 - P_H)(1 - P_H) \right].$$  

(27)

Then, substituting (28) into (27) yields

$$E\{H\} = \sum_{k=0}^{\infty} k \left\{ P_H^{(0)} \left[ (P_H(1-P_H))^k \right] \right\}$$

$$E\{H\} = P_H^{(0)} [P_H + (1 - P_H)] \sum_{k=0}^{\infty} k \left\{ (P_H(1-P_H))^k \right\}$$

$$E\{H\} = P_H^{(0)} \frac{P_H + (1 - P_H)^2}{(1-P_H)^2}$$  

(28)

where $P_H$ and $P_H^{(0)}$ are calculated by using (21). Finally, (28) can be used in (25) to obtain the handoff rate.

E. Call Forced-Termination Probability

Call forced-termination may result from either link unreliability or intercell handoff failure; but in general, a dropped call suffers $j$ ($j = 0, 1, 2, \ldots$) successful handoffs and one interruption (due to either a handoff failure or link unreliability) before it is forced to terminate. A successful handoff will occur with probability ($1 - P_h$). Thus, using the total probability theorem, the call forced-termination probability can be expressed as

$$P_{ft} = P_I^{(0)} + \sum_{j=1}^{\infty} P_I^{(0)} (1-P_h) [P_H(1-P_h)]^{j-1} P_I + P_I^{(0)} P_h$$

$$+ \sum_{k=1}^{\infty} P_H^{(0)} (1-P_h) [P_H(1-P_h)]^{k-1} P_H P_h$$  

(29)
where the call forced-termination is mathematically expressed in terms of the probabilities of handoff, handoff failure, and call-interruption due to link unreliability. Considering the channel holding time definition previously given, the condition for a call that is forced to terminate due to link unreliability is represented with the first set of summations on the right hand of (29); that is, with probability \( P_I \). The second set of summations on the right hand of (29) represents the case when a call is forced to terminate due to an intercell handoff failure, with probability \( P_h \).

In (29), the super index (0) in \( P_{HI} \) denotes the probability that a new call requires a handoff. Notice that it is necessary to differentiate between new and transferred calls because of the difference between the cell dwell time and the residual cell dwell time. Finally, (29) can be rewritten as

\[
P_{ft} = \left[ P_{f}^{(0)} + P_{I} P_{HI}^{(0)} \right] \frac{1 - P_h}{1 - P_{HI}(1 - P_h)} + P_{HI}^{(0)} \left[ P_h + \frac{1 - P_h}{1 - P_{HI}(1 - P_h)} \right].
\]

Expression (30) clearly suggests that the call forced-termination probability comprises two components. One is due to link unreliability, and the other is due to resource insufficiency. This new form for the call forced-termination probability reflects the impact of link unreliability in system performance. Notice that when the link unreliability is not considered \((P_I = 0)\), the call forced-termination probability given by (30) could be reduced to the call forced-termination probability found without considering the link unreliability (i.e., [40, eq. (35)]).

The proposed model here represents a general way to evaluate the system performance considering no particular call-admission control (CAC) policy. Particular expressions for new call blocking and handoff failure probabilities must be derived according to the specific CAC policy employed. This can be done by means of a quasi-birth and death analysis considering the interruption rate, as in [28] and [29]. Finally, the new call blocking and handoff failure probabilities can be obtained as the sum of the state probabilities in which new calls or handoff attempts are blocked, respectively.

VIII. NUmERICAL RESULTS

The goal of the numerical evaluations presented in this section is to understand and analyze the influence of link unreliability on system performance.

Unless otherwise specified, the following values of system parameters were used for the numerical results shown in this section: The offered traffic is 25 Erlangs/cell; the homogeneous system in which each cell has 30 total channels Blocked Calls Cleared (BCC) policy is considered to be CAC; the mean service time is \( E(X_s) = 180 \) s; and the mean of the cell dwell time is \( E(X_d) = 900 \) s. The cell dwell time and the unencumbered call-interruption time are modeled with \( n \)-order Erlang, exponential, or 2-order hyperexponential RVs, depending on the CV. For \( CV < 1 \) (\( = 1 \) \( > 1 \)), the Erlang (Exponential) \{hyperexponential\} distribution

is used. The considered values of the CV are \( CV = \{0, 1/\sqrt{9}, 1/\sqrt{8}, 1/\sqrt{6}, 1/\sqrt{5}, 1/\sqrt{4}, 1/\sqrt{3}, 1/\sqrt{2}, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}\). For the sake of comparison, two different values of the mean of the unencumbered call-interruption time are considered, that is, 1500 and 5000 s. Please note that both values of the mean of the unencumbered call-interruption time are significantly greater than the mean of the cell dwell time. The reason of this is that communication systems are commonly designed to be reliable; thus, the mean unencumbered call-interruption time should typically be greater than the mean service time and the mean cell dwell time. Different values of the mean unencumbered call-interruption time represent different reliability channel scenarios. Physically, it represents the probability that a call will be interrupted because of link unreliability. For instance, if the cell dwell time and the unencumbered interruption time are considered negative exponentially distributed, the probability that a call is interrupted due to link unreliability in a specific cell is \( P_I(E(X) = 1500) = 9\% \), whereas \( P_I(E(X) = 5000) = 2.9\% \). Thus, the channel is more reliable as the mean value of the interruption time becomes larger.

To validate the analytical model, simulations results are also obtained. Fig. 5 shows the forced-termination probability as a function of the CV of the cell dwell time obtained both analytically and by simulation. In Fig. 5, perfect agreement is shown between numerical and simulation results. Fig. 5 also reveals that the call forced-termination probability is an increasing function of the CV of the cell dwell time. Hence, the call forced-termination probability increases as the cell dwell times are more dispersed. This behavior can be explained as follows. If users stay inside cells for longer periods of time, the chances of a call to be dropped due to link unreliability is increased, and consequently, the call forced-termination probability increases (i.e., link unreliability becoming the main cause of call forced-termination). Figs. 6–8 show, in 3-D plots, respectively, the new call-blocking probability, call forced-termination probability, and carried traffic as a function of the CV of both cell dwell time.
Fig. 6. New call-blocking probability versus the CV of the cell dwell time and the CV of the unencumbered call-interruption time.

Fig. 7. Call forced-termination probability versus the CV of the cell dwell time and the CV of the unencumbered call-interruption time.

Fig. 8. Carried traffic versus the CV of the cell dwell time and of the unencumbered call-interruption time.

time and unencumbered call-interruption time. For values of the CV of the interruption time less than 1 (i.e., when the interruption time is Erlang distributed) and irrespective of the value of the CV of the dwell time, Figs. 6–8 show that there are no significant differences in the QoS metrics when the mean value of the interruption time changes from $E\{X_i\} = 1500$ s to $E\{X_i\} = 5000$ s. In other words, when the interruption time is Erlang distributed, the QoS metrics are not sensible to the mean value of the interruption time. From Fig. 7, it is observed that when the mean of the unencumbered call-interruption time is larger (i.e., $E\{X_i\} = 5000 \gg E\{X_i\} = 1500$), the call forced-termination probability is more sensitive to the CV of the cell dwell time. The reason is that as the mean value of the unencumbered call-interruption time decreases, fewer calls are forced to terminate due to link unreliability. Then, the main cause for call forced-termination is handoff failure (i.e., resource insufficiency).

On the other hand, when the unencumbered call-interruption time is considered to be hyperexponentially distributed (i.e., $CV > 1$), the numerical results show that there exists considerable differences among different cases regarding the CV of the hyperexponential distribution. Consequently, the performance metrics depend not only on the mean value of this RV but also on its variance. Notice also from Fig. 6 that the new call-blocking probability increases with the mean of the unencumbered call-interruption time. This behavior can be explained as follows. Smaller values of the mean unencumbered call-interruption time ($E\{X_i\} = 1500\text{ seg}$) result in smaller channel holding times (due to the interrupted calls), and more new calls can be served.

Fig. 7 shows that the call forced-termination probability is a decreasing function of the mean unencumbered call-interruption time. This is due to the fact that more calls are forced to terminate because of link unreliability. From Figs. 6–8, it is evident that the call forced-termination probability imposes severe penalties on the carried traffic.

Summarizing the foregoing results, we can say that, in general, the performance metrics are more sensitive to the CV of the different time variables when at least one of them is greater than one (i.e., hyperexponential distributed). In this case, the time variables are more dispersed, so the variations of the time variables are larger. Thus, as the time variables have a larger variance, the performance of the system decreases. In addition, notice that all of the performance metrics are more sensitive for variations in the CV of the unencumbered call-interruption time than that of the cell dwell time. This can be explained as follows. A call that needs a hand off (i.e., event with probability $P_H$) is forced to terminate only if there are no available resources in the target cell (i.e., with probability $P_h$). On the other hand, the unencumbered call-interruption time is directly forced to terminate call with probability $P_l$. So, it is concluded that the system performance is more sensible to link unreliability characterization than that of the users’ mobility.

IX. CONCLUSION

In this paper, an elegant and general analytical model for the system-level performance evaluation of mobile wireless networks taking into account both resource insufficiency and link unreliability is proposed and mathematically analyzed. In addition, a comprehensive overview of the studies on link
unreliability and nonexponential assumptions in mobile cellular communications networks is provided. The main contribution of this paper is the introduction of an interruption process and its use to achieve a general analytical model for the system-level performance evaluation of mobile wireless communication networks considering both resource insufficiency and link unreliability. This interruption process has a potential associated time that is called “unencumbered call-interruption time.” It was shown that the probability distribution of this time can be derived in terms of measurable system-level parameters that can easily be obtained at BSs in real cellular networks. From link-level statistics, it has been shown that the unencumbered call-interruption time should be modeled as a generally distributed RV to capture the overall effect of the channel. More importantly, novel mathematical expressions for call forced-termination probability, handoff probability, and handoff call arrival rate that capture both resource insufficiency and link unreliability were derived. Numerical results show that the developed model allows us to obtain new and important insights into the dependence of system performance on link unreliability; specifically, it could be concluded that the system performance is more sensible to the interruption process than to the mobility. The proposed mathematical model is very general and includes previously proposed models in the literature as particular cases. For instance, when the mean unencumbered call-interruption time is considered to be infinite, the proposed model is reduced to the well-known model when no link unreliability is considered.

Finally, this paper has called again for the necessity of reexamining classical analytical results in traffic theory, which are used for the analysis and design of wireless mobile networks that have not yet considered link unreliability. This includes mobile wireless communication systems with different radio resource-management strategies, multiple services and/or traffic classes, diversity schemes, etc. Future work can include the generalization of the mathematical model for considering unencumbered service (i.e., call holding) time with phase-type distributions. When phase-type probability distributions are considered, the proposed mathematical model allows elegant teletraffic analysis for the performance evaluation of mobile wireless networks by using quasi-birth and death processes. Additionally, to accurately characterize link unreliability, it is also important to find the distribution of the unencumbered call-interruption time directly from the channel model by considering a more general channel model such a Rayleigh/Nakagami. All these issues are subjects of further future research.

References


