Impact of Signaling Load on the UMTS Call Blocking/Dropping

Saowaphak Sasanus and David Tipper
Telecommunications Program, School of Information Sciences
University of Pittsburgh
135 N. Bellefield Avenue, Pittsburgh, PA 15260
Email: sasst128, dtipper@pitt.edu

Yi Qian
National Institute of Standards and Technology
100 Bureau Drive, Stop 8920, Gaithersburg, MD, 20899
Email: yqian@nist.gov

***Abstract*— Radio resources in the third generation (3G) wireless cellular networks (WCNs) such as the universal mobile telecommunications system (UMTS) network is limited in term of soft capacity. The quality of a signaling service transmission depends on various factors (i.e., location, speed, and the environment of the user, power control, and variable data rate requirement), and has impact on quality of user data communications where the opposite order is also true. In this paper, we provide the first step to evaluate the impact that various signaling service types have on call blocking and ongoing call drop in the UMTS systems. The radio resource's acquisition time for each signaling service type is calculated. The maximum number of sessions that a signaling service type can be transmitted simultaneously is estimated along with the converting value when the other signaling service type is transmitted instead. Our analysis reduces the computational complexity in the call admission control (CAC) and allows the preservation on classes of services. An example of traffic scenario along with the analytical result are given illustrating the benefit of our study.**

I. INTRODUCTION

To increase available capacity, the 3G WCNs adopt wideband code division multiple access (WCDMA) technology where user data and signaling services are transmitted over the same large frequency range. A transmission of a user data traffic can be distinguished from a transmission of a signaling traffic through orthogonal codes. However, due to the limitation of the orthogonal codes and the code allocation algorithm [1], interference becomes the limit to the radio capacity. In the UMTS networks that uses the frequency division duplex mode, two common types of the interference are inter-and intra-band interference, and inter-and intra-cell interference. Here, we consider only the later.

An increase in the variety and the heterogeneity of signaling traffic obviously degrades quality of user data communications, and vice versa. Thus, radio resource must be carefully allocated in order to preserve quality of service (QoS) in signaling and user data traffic. Unfortunately, the radio resources is not the only scarce signaling resource in the cellular networks. Database servers are also required to support seamless roaming and secure communications. Thus, to ensure quality of signaling services, the mechanisms include a CAC and a signaling overload control must both be in place to maintain quality of voice and multimedia calls. For global system for mobile communications (GSM) networks, we proposed a signaling overload control algorithm that considers the scarcity of both radio and database server in [2] [3]. Here, we develop the material that can be used to construct the similar set of algorithms with the specific attention for the UMTS networks.

In the current literature, only a few simulation based studies have happened on the impact of some signaling services (i.e, location update, paging) on user data communications [4] [5]. In this work, we illustrate the impact of most fundamental signaling services on the cellular communications (e.g., call setup, location update, and handoff). The available radio resources is represented by the numbers of sessions that each of these signaling service called the saturation rate can be simultaneously supported. The saturation rate is calculated from the acquisition time that each signaling service needs to utilize the orthogonal codes in up-link and down-link. Code holding time can be derived from the transmission rate of the air interface with a choice of either common or dedicated control channel (CCH or DCH) and the signaling message length gathering from the signaling procedures discussed in [6] [7]. We also develop an simple equation that allows a conversion between the saturated number of sessions of one signaling service type to that of the other service type based on a well known signal-to-interference ratio (SIR) formula [4] [8]. By the realization of data traffic demand causing from the acceptance of various signaling service types, we can compare the impact that one signaling service creates to that of the others.

By knowing this information, we can efficiently plan radio resource allocation for various classes of signaling services. We later apply our findings with the CAC of the UMTS system integrated with signaling overload control at the database servers, preserving classes of services at both air and the database servers. The remainder of the paper is organized as follows. In Section II, the literature on CAC is briefly reviewed to illustrate the need of the saturated rate estimation and the basis of the conversion number (i.e., based on the SIR constraint). Also, the message flow of each signaling procedure along with its length is depicted. In Section III, the acquisition time is calculated. Followed by the approximation of the saturation rate that each signaling service type can be transmitted simultaneously in the air interface within the control interval. The analytical model is given for a conversion between saturation rate of one signaling service type to the other. The benefit of our analysis is illustrated by an example of traffic scenario along with the analytical result in Section IV before we state our conclusions.

II. LITERATURE REVIEW

A. Call Admission Control

A CAC algorithm accepts or rejects the arrival service requests based on the current system status. We address the existing CAC algorithtms in three perspectives. First is the method to reject new calls [9] [10]. For example, complete sharing allows all classes of signaling services to share the same pool of the available radio resources, whereas the threshold-based CAC restricts services from the lower classes by using multiple thresholds.

Second is the parameter that represents the status of radio resources (e.g., the interference, the received signal power, the SIR). The new calls are only accepted if the maximum or minimum of the parameter are not violated. The SIR-based CAC more accurately estimates the current system status compared to the interference-based and the power-based CACs since it can differentiate between the received signal power and the interference.

Third is the method to find the available radio resources in terms of a representative parameter. For example, the interference of mobiles within the same cell may be used to estimate the number of sessions that the available radio resources sufficiently serve, or the interference of mobiles from the other cells may also be included into the estimation. However, the representative parameter are unnecessary in some CACs that directly apply the parameter into the rejection method. For example, a CAC that accepts a new call after a test pilot. The SIR measured within the test pilot is compared with the minimum SIR to decide whether to accept or reject the call.

In this work, we propose a SIR-based CAC that pre-calculates the maximum number of signaling sessions that the current available radio resources can support within the next control interval. This information allows efficient allocation of the radio resource, maintaining class of service. We calculate the maximum number of sessions that each of various signaling services can be simultaneously served, and provide the converting values among them.

B. Signaling Procedures

The following signaling services are studied: new call request, paging, location update, handover, SMS, and end call requests. We illustrate the procedures of these signaling services through their message flows. The signaling message length is also given for the purpose of the calculation in the code acquisition time.

First, we consider the signaling services that effect the quality of the active user-data transmission on the up-link direction (i.e., location update, call setup, handover$_{org}$, and SMS$_{org}$). The user equipment (UE) must perform a general packet radio service (GPRS) attach, the security related procedures, and the packet data protocol (PDP) context before sending the data if any. The GPRS attach allows the system to handle the mobility management and to obtain the detailed location information. The PDP context characterizes sessions and assigns the PDP address for each PDP session. These procedures are illustrated in Figure 1.

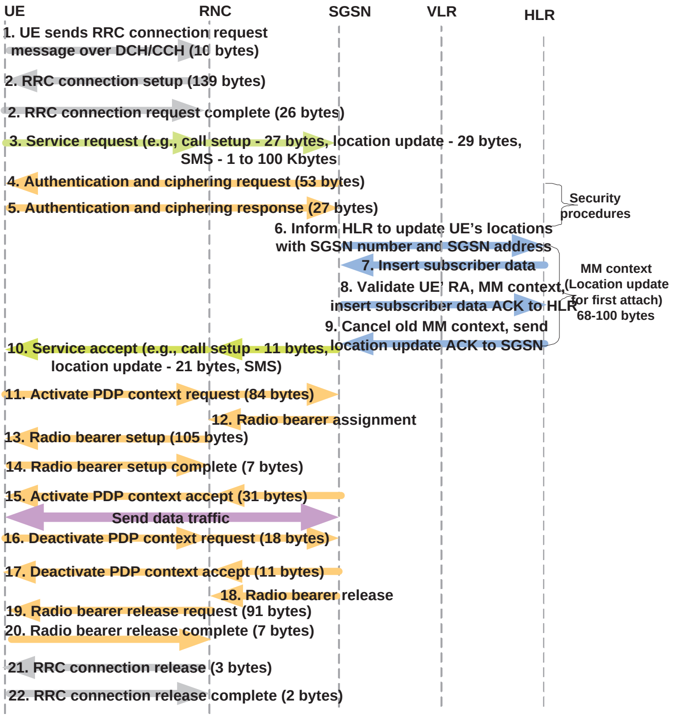

Fig. 1. The GPRS attach and a PDP context [11]

According to [7], these signaling procedures consist of the following steps. In step 1, the radio resource control (RRC) connection is established over the CCH. Then, in step 2, the radio network controller (RNC) sets up a point-to-point radio connection as well as the signaling connection to the network before sending acknowledgment back to the UE. After that, the UE will start the attach process in steps $3-10$ which includes the attach request, the identity request/response for the first time that the UE is attached to the network, the authentication request/response if the mobility management context does not exist for the UE anywhere else. Then, the PDP context will be setup to characterize the radio bearer (RAB) session and RAB request is setup in step $11-15$. The PDP addresses that will be used and stored at the UE and the GPRS supported nodes (GSNs) are activated. The PDP context contains mapping and routing information for packet transmission between the UE and the gateway GSN (GGSN). After the UE finished data transmission, the RAB release is initiated along with the PDP context deactivation and the RRC release in step $16-22$.

Second, we consider the signaling services that interfere with the user data communications in the down-link direction (i.e., paging, handover$_{org}$, and SMS$_{term}$). Sometimes, a SMS$_{term}$ also needs the paging service if the terminating UE is in the idle mode. We study the effect of paging to system capacity of each cell. In the UMTS network, the user locations are tracked in terms of the location area (LA) for the circuit-switched domain and the routing area (RA) for the packet-switched domain. A LA consists of multiple RAs. In turn, each RA consists of multiple UTRAN registration areas (URAs) each of which consists of multiple cells.

In the packet-switched domain, the UE stays in the idle mode when a UE does not establish any connection. The UE locations are tracked with the accuracy in the level of RA. The UE state is moved to cell-connected only when the connection is established. If later the UE is inactive longer than timeout, the UE state is moved to the URA connected and the tracking accuracy is in the level of URA. If the terminating UE is not in the RRC cell-connected state, the HLR will be queried for the availability, the billing information, the available services, and the last known LA or RA of the UE. Then, the core network pages all cells within the UE's LA or RA over the paging channel (PCH). The larger the location area, the larger the paging but the smaller load of the location update. After that, the UE sends the response to the BS in the random access control channel (RACH), which triggers the BS to assign the traffic channel to the UE. Then, the RRC connection is established following with the delivery of the SMS message (for SMS service).

We note here that PCH, RACH, and another forward access control channel (FACH) is later referred to as CCH.

III. ANALYTICAL ANALYSIS

A. The Acquisition Time

Most of signaling services can be delivered over either the CCH or the DCH, leading to the different code acquisition time. The CCH benefits from fast transmission since it does not require call setup or tear-down, and the ability to share code. Also, the interference is introduced only when the signaling services is transmitted, not in the idle period unlike in DCH. However, it lacks of fast power control which anticipates higher interference than CCH. On the other hand, the DCH allows fast power control, but the interference is always generated even when channel is idle.

According to the study in [12], the CCH is more suitable to lower burst size compared to the DCH. More specifically, the CCH performs better than DCH for a signaling service session which transmits signaling messages of size approximately up to 250 bytes. Because the CCH access time is shorter than the setup time of DCH. In the up-link, the maximum data rate for the CCH and DCH are 60 kbps and 48 kbps for a spreading factor of 32. In the down-link, the CCH and DCH can accommodate the maximum transport channel rate of 36 kbps and 28.8 kbps for a spreading factor of 64.

Table I summarizes the acquisition time which can be derived from the total message length according to [13], and the channel data rate. LU considered here is the periodic LU where GPRS attach and security command are not performed. We use the maximum length of SMS message, 1Kbytes.

TABLE I
THE CHANNEL ACQUISITION TIME

Service type	MSG length (bytes)		Acquisition time (ms)	
	DCH	CCH	DCH	CCH
SMS	1180	1000	204.4	133.3
Location update	394	214	81.6	38.6
Call setup	652	472	148.9	88.9
End call	689	500	155.3	93.8
Paging	-	9	-	2.0
Inter-RNC Handoff	-	17	-	2.71
UE offline	199	45	37.7	36.6

B. The Maximum Signaling Service Sessions

In this section, we roughly estimate the maximum amount of the signaling service sessions that can be conveyed by mean of a SIR analysis, based on the basic equation adopt from [8]. By assuming negligible the interference and noise and the equal received signal power from all users, signal to noise ratio (SNR) is $\frac{S}{(N-1)S}$ where N is the total number of users in the cell and S denotes the received signal power. SIR which is energy-per-bit to noise power spectral density is $\frac{S/R}{(N-1)S/W}$ where W is the total radio frequency bandwidth, and R is the baseband information bit rate.

In this work, we consider arrivals within each control interval. We assume that only the signaling service type i is initiated at the beginning of the control interval time between $t-1$ to t. Let S^P be the received signal power of the active signaling services initiated within the previous control interval measured at time t which concerns the period of time before $t-1$. The requirement of the SIR for a signaling service type i, SIR$_i$ can be calculated as shown in Eq. 1. Note here that our analysis here is also applicable for data traffic.

$$\begin{aligned} SIR_i &= \frac{S_i/R_i}{((1-\alpha)I_{in}+S_{out}+N_0)/W} \\ &= \left(\frac{W}{R_i}\right)\frac{S_i}{R_i(1-\alpha)[S^P+(N_i-1)]S_i+S_{out}+N_0} \qquad (1) \end{aligned}$$

α is the orthogonal factor in the down-link and the interference reduction scheme in the up-link. There is no synchronization among users in the up-link, so there is no orthogonality. We assume that the transmission in one direction have no impact to the data rate in the other direction. Only intra-cell and inter-cell interference is included in the calculation. I_{in} and S_{out} are defined as the interference caused by transmission of other services within the same cell and within the other cells, respectively. In fact, I_{in} is only S^P, and S_{out} is the summation of I_{in} from the neighbor cells. N_i denotes the maximum number of sessions that signaling service type i can be supported simultaneously by the available radio resources within the control interval. R_i be the baseband information bit rate of the signaling service S_i.

The BS can simply monitor the received signal power for an analysis of the up-link transmission. For the down-link, the received signal power is calculated according to [14] derived from the transmitted signal power in Table II and the path loss model adopted from [15], $S = P_t - max(P_l - G, C_l)$. S and P_t are the received and transmitted power in dBm. G denotes the antenna gain in the BS (11dB), and C_l is the maximum coupling loss (70dB). The path loss denoted by P_l is $128.1 + 37.6logr$ in dB where r is the distance between the UE and the BS in km.

In the interference limit system such as the UMTS, noise is negligible compared to the interference, $N_0 \rightarrow 0$. We can find N_i as follows.

$$\begin{aligned}
N_i &= \frac{W}{R_i(1-\alpha)SIR_i} - \frac{S^P}{S_i} - \frac{S_{out}}{(1-\alpha)S_i} - \frac{N_0}{R_iS_i} + 1 \\
N_i &= \frac{W}{R_i(1-\alpha)SIR_i} - \frac{S^P}{S_i} - \frac{S_{out}}{(1-\alpha)S_i} + 1 \\
N_i &= \frac{a}{R_i} - \frac{b}{S_i} - \frac{c}{S_i} + 1 \qquad (2) \\
&\text{where : } a = \frac{w}{(1-\alpha)SIR_i},\ b = S^P,\ c = \frac{S_{out}}{1-\alpha}
\end{aligned}$$

Let N_i be the maximum number of sessions that signaling service type i alone can be supported by the available radio resources within the control interval, and V_{ij} indicates the value that converts N_i to N_j.

$$\begin{aligned}
N_j &= V_{ij}N_i \\
\frac{a}{Rj} - \frac{b}{S_j} - \frac{c}{S_j} + 1 &= V_{ij}\left(\frac{a}{Rj} - \frac{b}{S_j} - \frac{c}{S_j} + 1\right) \\
V_{ij} &= \left(\frac{R_i}{R_j}\right)\left(\frac{S_i}{S_j}\right)\left(\frac{F_j}{F_i}\right) \qquad (3) \\
where : \quad & F_j = aS_j - (b+c)R_j + 1 \\
& F_i = aS_i - (b+c)R_i + 1
\end{aligned}$$

Let assume that only S_i and S_j exists over the control interval. From the total available number of sessions N_i, the followings are derived for the case that X sessions are used by S_i and $N_i - X$ sessions of S_i are occupied by S_j. Let denote the number of sessions that S_j can be supported by $N_i - X$ sessions of S_i by $\acute{N}_j$. The conversion value $\acute{V}_{ij}$ which maps the number that signaling service type S_i can be supported by the available radio resource to the number that S_j can be supported is shown in Eq. 4.

$$SIR_j = \frac{\frac{S_j}{(1-\alpha)(S^P + XS_i + ((N_i - X)Var_{ij} - 1)s_j) + S_{out} + N_0}}{R_j/W} \qquad (4)$$

$$\begin{aligned}
\acute{N}_j &= V_{ij}(N_i - X) = \frac{a}{R_j} - \frac{b + c + XS_i}{S_j} + 1 \\
\text{where : } \quad & a = \frac{w}{(1-\alpha)SIR_i},\ b = S^P,\ c = \frac{S_{out}}{1-\alpha}
\end{aligned}$$

$$\begin{aligned}
\acute{V}_{ij} &= \frac{\acute{N}_j}{N_i} = \frac{\frac{a}{R_j} - \frac{b+c+XS_i}{S_j} + 1}{\frac{a}{R_j} - \frac{b+c}{S_j} - X + 1} \\
&= \left(\frac{R_j}{R_i}\right)\left(\frac{S_j}{S_i}\right)\left(\frac{F_i - XR_iS_i}{F_j - XR_jS_i}\right) \qquad (5) \\
where : \quad & F_i = aS_i - (b+c)R_i + 1 \\
& F_j = aS_j - (b+c)R_j + 1
\end{aligned}$$

By using the induction method, Eq. 5 becomes Eq. 3 when $X = 0$. With the similar assumption above, Eq. 5 below is the general form of V_{ij} where $X_1, X_2, ...X_{T_y}$ signaling service sessions of $S_1, S_2, ..., S_{T_y}$ is transmitted over the control interval for the total of T_y signaling service types.

$$\begin{aligned}
V_{ij} &= \left(\frac{R_j}{R_i}\right)\left(\frac{S_j}{S_i}\right)\left(\frac{F_i - f_i(T_y)}{F_j - f_j(T_y)}\right) \qquad (6) \\
where : F_i &= aS_i - (b+c)R_i + 1 \\
F_j &= aS_j - (b+c)R_j + 1 \\
f_i(T_y) &= R_iS_i(X_1 + ... + X_{j-1} + X_{j+1} + ... - X_{T_y}) \\
f_j(T_y) &= R_jS_i(X_1 + ... + X_{j-1} + X_{j+1} + ... - X_{T_y})
\end{aligned}$$

From the analysis results, we can promptly plan types of signaling services and its amount that will be accepted based on its class at the beginning of the control interval despite large signaling service types in the near future. At every control interval (e.g., $1s$ for signaling services), the computation complexity is reduced from $O(T_y^2)$to $O(T_y)$ where T_y is the number of signaling service type. For $O(T_y^2)$, all $N_1, N_2, ..., N_{T_y}$ must be calculated first before the calculation of $V_{12}, V_{13}, ..., V_{1T_y}$. Whereas, for $O(T_y)$, only N_1 and $V_{12}, V_{13}, ..., V_{1T_y}$ are needed. Signaling service that is most frequently occurred (e.g., location update) should be assigned as the signaling service type 1, so the estimation of the saturated rate or the max. number of sessions can be more accurate.

The actual usage of the radio resources can be very different from the radio resource allocation plan, as user's characteristics (e.g., environment, mobility, and interference) changes over times, especially in large control interval. Thus, within the control interval, we should adjust radio resource pool and allocation according to the current user's status (e.g., every $0.33s$ from the total of 1s control interval). The adjustment period can be adaptively set according to change in the user's status. S^P becomes the received signal power of services within the previous control interval and the signaling services that are already admitted within the current control interval in Eq. 1. Because of this adaptability need, using our formulation will further reduces the computation complexity in the admission control.

IV. ANALYTICAL RESULTS

We use the example scenario when user either connects with low speed data 12.2 kbps or high speed data session 64 kbps after call setup or handoff to new cell. The data rate for CCH and DCH are set as calculation in the Table I. Other parameters are set as shown in Table II.A. From both tables, we derive the maximum number of sessions for various channel rate at the beginning of the control interval in Table II.B. Low and High indicates low and high speed data channel. Since the capacity is limited only by load in the down-link, we perform here only an analysis for down-link with an assumption that load in the down-link is higher than that in the up-link. Load in up-link is only influenced the coverage.

TABLE II

(A) Power Control Parameters (B) Max. No. of Sessions

User data parameters	PS
Bit rate(kbps)	12.2 (LOW), 64 (HI)
Spreading gain	32 (UL), 64 (DL)
SIR requirement(dB)	2.5 [4]
BS transmitting power(W)	20 (DCH), 3 (CCH)
Orthogonal factor	0.5
Activity factor	1
Control interval (sec)	1

Ch. Type	max.no.of sessions
CCH	70
DCH	101
Low	305
High	183

In the analysis, only one session of data traffic is initiated for call setup and handoff. Table III shows the maximum number of sessions for some fundamental signaling services available within the control interval $1s$.

TABLE III

The Maximum Number of Signaling Service Sessions (Over 1s)

Signaling Type	Max. no of sessions	
	CCH	DCH
SMS	756	346
Location update	2612	868
Call setup	219 (Low), 179 (High)	213 (Low), 179 (High)
End call	1134	476
Paging	-	50405
Inter-RNC Handoff	301 (Low), 183(High)	- (Low), - (High)
UE offine	1878	2754

We illustrate the benefit of our analysis through the a small network consisting of one node B with the arrival signaling traffic load in the Table IV. Here, we compare between two cases: a simple CAC which is equipped and not equipped with the knowledge of the estimated saturated rate in advanced. The CAC rejects the arrival traffic only if there is no available radio resource. The table shows that classes of services can be improved as our analysis is included into the simple CAC.

V. Conclusions and the Future Work

In this paper, we provided a simple analysis to study the impact of signaling load on call blocking/dropping. The calculation of the data rate requirement for basic signaling services are given in both common and dedicated radio channels with our suggested selection. Based on the estimation of the maximum number of users for soft capacity system discussed in [8], we calculate the simple conversion that allows fast mapping between the maximum number of signaling service sessions of one signaling service type to the other.

Our ongoing work applies the findings of this work to the CAC that is aware of the database servers's available resource for the UMTS networks.

TABLE IV

The Benefit of our Derivation

Traffic load (class)	Avail. radio resource pool	Equipped		Non-equipped	
		Served Traffic	Rejected Traffic	Served Traffic	Rejected Traffic
SMS (LOW)	XX	XX	XX	XX	XX
LU(MED)	XX	XX	XX	XX	XX
Call setup(LOW)	XX	XX	XX	XX	XX
Paging(MED)	XX	XX	XX	XX	XX
Handoff (HI)	XX	XX	XX	XX	XX

Acknowledgments

This work was supported by a scholarship from TOT Corporation Public Co. ltd. The author would like to thank Dr. P. Krishnamurthy for his advises.

References

[1] T. Al-Meshhadany and K. A. Agha, “A new code allocation scheme for UMTS system,” in *IEEE Vehicular Technology Conference, 2001. VTC 2001)*, 2001.

[2] S. Sasanus and D. Tipper, “On the level of guaranteed services for signaling control in cellular networks,” in *IEEE Global Communications Conference (GC’07)*, 2007.

[3] ——, “Adaptive multi-class signaling control for cellular networks,” in *IEEE Symposium on Computers and Communications (ISCC’07)*, 2007.

[4] D. Lin, B. Yeo, and Y. C. Y. Kwok, “Effects of location management signaling load on the forward link throughput of UMTS-FDD systems,” in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2005)*, Sep. 2005.

[5] K. Thum, B. Yeo, Y. Chew, and K. Ang, “Performance study of the varying parameters on the paging and updating signaling loads in an UMTS-FDD system,” in *Global Telecommunications Conference, 2004. (GLOBECOM’ 04)*, Dec. 2004.

[6] Y.-B. Lin, A.-C. Pang, Y.-R. Haung, and I. Chlamtac, “Mobility management: from GPRS to UMTS,” *Wireless Communications and Mobile Computing*, vol. 1, no. 4, pp. 339–359, Aug 2001.

[7] A. Mishra, “Performance characterization of signaling traffic in UMTS core networks,” in *IEEE Global Telecommunications Conference, 2003.(GLOBECOM ’03)*, 2003.

[8] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, New Jersey: Prentice Hall, 2001.

[9] M. Andersin, Z. Rosberg, and J. Zander, “Soft and safe admission control in cellular networks,” in *IEEE/ACM Transaction on Networking*, vol. 5, no. 2, Apr. 1997, pp. 255–265.

[10] D. Kim, “Efficient interactive call admission control in power-controlled mobile systems,” *IEEE Transactions on Vehicular Technology*, vol. 49, no. 3, pp. 1017–1028, May 2000.

[11] ETSI, *TR 101 503 v8.27.0. Digital Cellular Telecommunications System (Phase 2+); Mobile Radio Interface Layer 3 Specification; Radio Resource Controll (RRC) Protocol.*

[12] J. Sachs, S. Wager, and H. Wiemann, “Performance of shared and dedicated resources in WCDMA,” in *IEEE Wireless Communications and Networking Conference, 2000. (WCNC’00)*, 2000.

[13] R. Kreher and T. Ruedebusch, *UMTS Signaling: UMTS Interfaces, Protocols, Message Flows and Procedures Analyzed and Explained*, 2nd ed. 111 River Street, Hoboken, NJ 07030, USA: John Wiley & Sons, Ltd., 2005.

[14] 3GPP, *TR 25.942 v3.3.0. RF System Scenarios.*

[15] ETSI, *TR 101 112 v3.2.0. Selection procedures for the choice of radio transmission technologies of the UMTS.*