Performance of CS Fallback for Long Term Evolution Mobile Network

Ren-Huang Liou, and Yi-Bing Lin, *Fellow, IEEE*
Department of Computer Science
National Chiao Tung University
{rhliou, liny}@cs.nctu.edu.tw

Yingrong Coral Sung
Center for Information and Communications Technology
National Chiao Tung University
ycsung@cs.nctu.edu.tw

Pang-Cheng Liu
Department of Computer Science
National Chiao Tung University
pcliu@cs.nctu.edu.tw

Christian Wietfeld
Communication Networks Institute
TU Dortmund University
Christian.Wietfeld@tu-dortmund.de

Abstract

When a mobile operator migrates its network from the *3rd Generation* (3G) system to *Long Term Evolution* (LTE), both 3G and LTE will co-exist for a period of time. Since the 3G *Circuit-Switched* (CS) voice mechanism is more mature and available than that for LTE *Voice over Internet Protocol* (VoIP), the operator may consider CS fallback as a solution to provide reliable voice calls. According to the *3rd Generation Partnership Project* (3GPP) CS fallback procedure, when a mobile user in the LTE network has an incoming or an outgoing call, the *User Equipment* (UE) falls back from LTE to *Universal Mobile Telecommunications System* (UMTS). When the call is complete and released, the UE immediately returns to LTE. If the next activity for the UE is another voice call, immediately switching from UMTS to LTE may not be efficient. In this case, the UE has to perform another CS fallback. To resolve this issue, we suggest delaying the returns to avoid unnecessary CS fallbacks, which is called *delayed-return* (DR). Based on the measurements from the real UMTS and LTE networks, we develop analytic model to investigate the performance of the CS fallback with DR. The study indicates that the DR scheme can effectively reduce the CS fallback costs up to 60%.

Index Terms: CS fallback, delayed-return scheme (DR), long term evolution (LTE)

1 Introduction

The *3rd Generation Partnership Project* (3GPP) *Long Term Evolution* (LTE) [1] defines an all IP network architecture that evolves from the *Universal Mobile Telecommunications System* (UMTS) [2] to provide high-speed data services. When a mobile operator migrates its network from the *3rd Generation* (3G) system to LTE, both 3G and LTE will co-exist for a period of time, and initially, the 3G coverage is more complete than the LTE coverage. Since LTE is a *Packet-Switched* (PS) network, the *Circuit-Switched* (CS) voice service [3] is not supported as 3G does. Therefore the LTE voice service is offered through the *Voice over Internet Protocol* (VoIP) technology. Since the 3G CS voice mechanism is more mature and available than that for LTE VoIP, in the deployment of LTE, many operators consider CS fallback [4] as the solution to provide reliable voice calls. The CS fallback technique switches the *User Equipment* (UE; the mobile phone) from the LTE network to the 3G legacy system when a voice call is attempted. In order to use the CS domain in the UMTS system when the UE resides in LTE, the LTE network needs to register the UE with both the LTE and the UMTS CS domain and delivers the CS paging message from UMTS to the UE.

Figure 1 illustrates a simplified architecture of UMTS and *Evolved Packet System* (EPS) for LTE. This architecture includes two parts: the UMTS network and the LTE network.

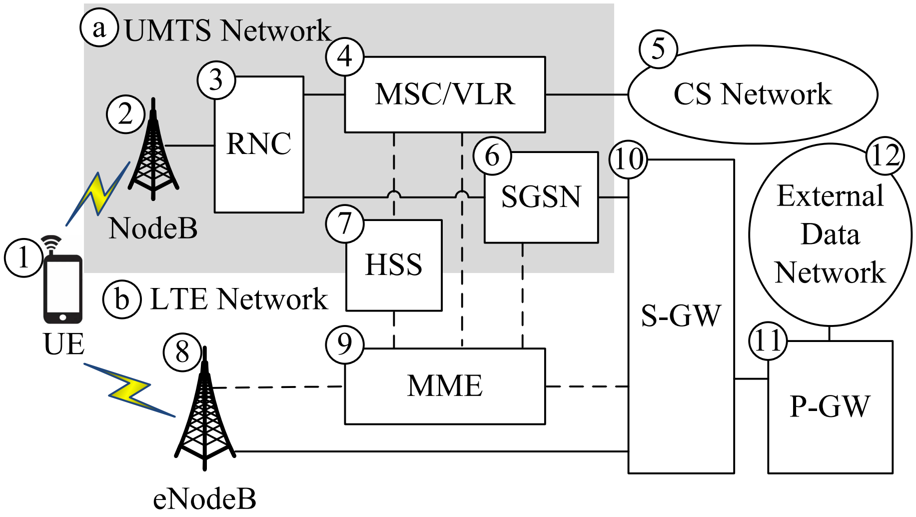


Figure 1: The EPS and UMTS Architecture for CS Fallback (dashed lines: signaling; solid lines: signaling/data)

A UE (Figure 1 (1)) accesses UMTS and LTE services through the radio interfaces. In the UMTS network (Figure 1 (a)), the *UMTS Terrestrial Radio Access Network* (UTRAN) consists of NodeBs (Figure 1 (2)) and *Radio Network Controllers* (RNCs; Figure 1 (3)). A NodeB provides *Wideband Code Division Multiple Access* (WCDMA) radio connectivity between the UE and the corresponding RNC. The RNC connects to the UMTS core network. This core network is partitioned into the CS and the PS domains. The CS domain includes *Mobile Switching Centers* (MSCs) and *Visitor Location Registers* (VLRs; Figure 1 (4)). An MSC is responsible for call control and connection between the UE and the external CS Network (Figure 1 (5)). A VLR is responsible for the mobility activities of the MSC. The PS domain consists of *Serving GPRS Support Nodes* (SGSNs; Figure 1 (6)) which provide the mobility and session services to the UEs. The *Home Subscriber Server* (HSS; Figure 1 (7)) is the master database containing all user-related subscription information, which supports mobility management of mobile users. In the LTE network (Figure 1 (b)), the *Evolved UMTS Terrestrial Radio Access Network* (E-UTRAN) consists of *evolved NodeBs* (eNodeBs; Figure 1 (8)) to offer LTE radio connectivity to the UE. The E-UTRAN connects to the LTE core network that includes the following components. A *Mobility Management*

Entity (MME; Figure 1 (9)) interacts with the HSS to offer mobility management and session control. The *Serving Gateway* (S-GW; Figure 1 (10)) is responsible for routing data packets and is an anchor of the user plane data for intra- and inter-system handovers. The *Packet Data Network Gateway* (P-GW; Figure 1 (11)) provides the connectivity to the External Data Network (Figure 1 (12)) and the per-user based packet filtering. In the UMTS PS domain, the SGSNs connect to the External Data Network through the S-GW and the P-GW. According to the 3GPP CS fallback procedure [4], when a mobile user in the LTE network has an incoming or an outgoing call, the UE falls back from LTE to UMTS. When the call is complete and released, the UE immediately returns to LTE. If the next activity for the UE is another voice call, immediately switching from UMTS to LTE may not be efficient. In this case, the UE has to perform another CS fallback. To resolve this issue, we suggest delaying the returns to avoid unnecessary CS fallbacks. This paper is organized as follows. Section 2 describes the CS fallback procedures, the existing *Immediate-Return* (IR) scheme and the proposed *Delayed-Return* (DR) scheme for returning to LTE. Section 3 proposes an analytic model for the IR and the DR schemes. Section 4 studies the performance of IR and DR by numerical examples, and conclusions are given in Section 5.

2 3GPP CS Fallback Procedures

This section describes the CS fallback procedures defined in the 3GPP, including call setup and call release with IR. We also report the measured processing times for the procedures collected in live 3G and LTE networks in [5, 6, 7]. Then we introduce the DR scheme, including call release and data session setup.

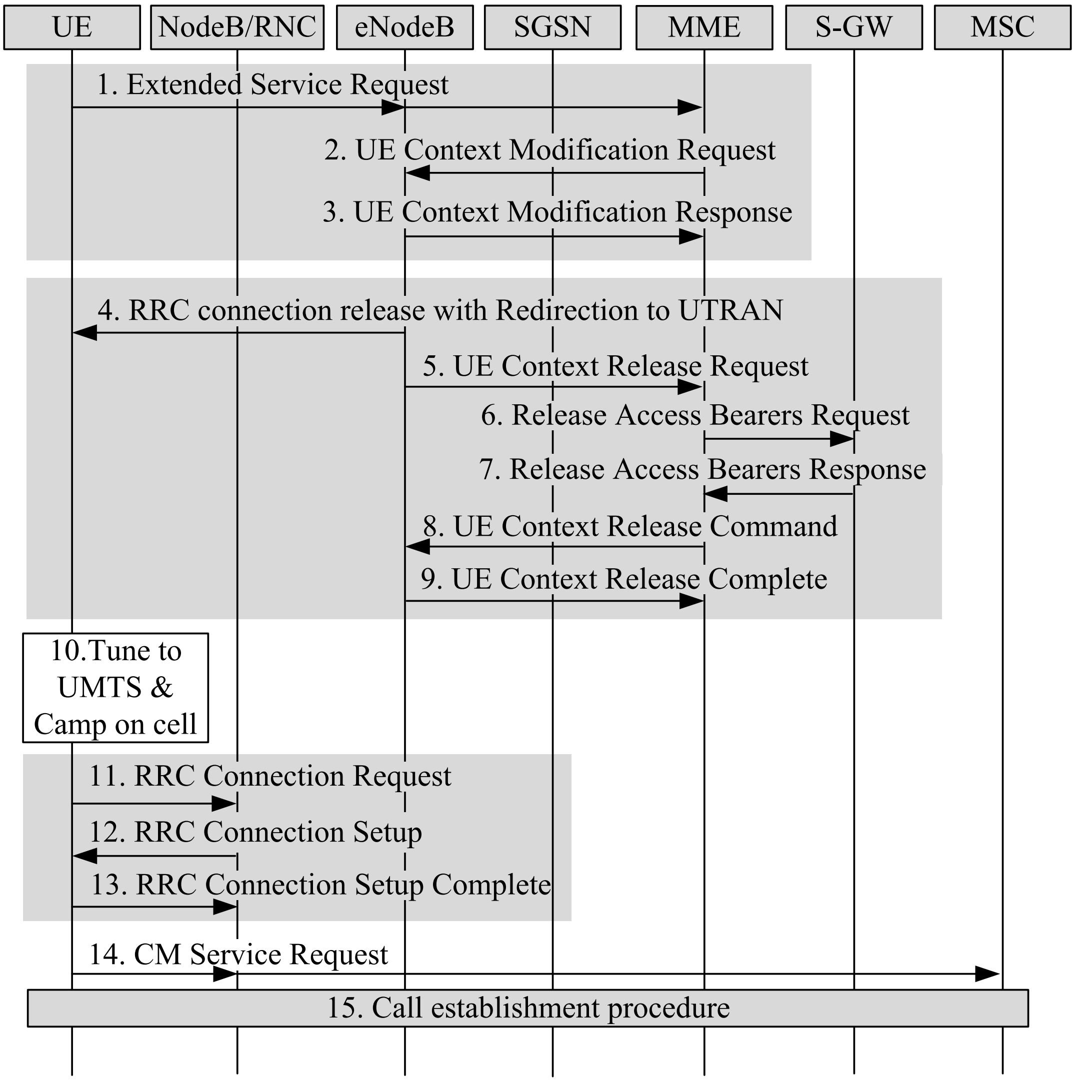


Figure 2: Call Setup with CS Fallback

2.1 LTE Call Setup with CS Fallback

Figure 2 illustrates the CS fallback message flow when a UE makes a call in the LTE network. The following steps are executed:

Step 1. The UE sends the Extended Service Request message to the MME to initiate the CS fallback procedure.

Steps 2 and 3. The MME exchanges the UE Context Modification Request and Response message pair with the eNodeB to indicate that the UE should fall back to the UTRAN. Note that Steps 1-3 take about 0.3 seconds [5].

Step 4. The eNodeB sends the UE the *Radio Resource Control* (RRC) Connection Release with Redirection to UTRAN message to indicate that it may follow the cell identity and System Information to attach to the corresponding UTRAN cell.

Steps 5-9. Parallel to Step 4, the eNodeB sends the UE Context Release Request message to the MME to release the bearers between the eNodeB and the S-GW. Steps 4-9 take about 0.2 seconds [5].

Step 10. After Step 4, the UE tunes the radio to UMTS, and camps on the NodeB according to the System Information in the RRC Connection Release with Redirection to UTRAN message. Step 10 takes about 2.3 seconds for 3GPP R8 and 0.3 seconds for 3GPP R9 [5].

Steps 11-13. The UE exchanges with the NodeB the RRC Connection Request and Setup message pair to establish the radio connection. Then the UE sends the NodeB the RRC Connection Setup Complete message to acknowledge the RRC connection establishment. Steps 11-13 take about 0.3 seconds [5].

Steps 14 and 15. The UE sends the *Call Management* (CM) Service Request message to initiate the CS call establishment procedure. The UE includes the *Circuit-Switched Mobile Originated* (CSMO) flag to indicate that it is a CS fallback call. The CS call establishment at Step 15 follows the 3GPP standard, and the details can be found in [8]. Steps 14 and 15 take about 3.5 seconds [5].

Note that if the UE in LTE is engaged in a data session when a call arrives, then the PS connection (for the data session) is also switched to UMTS in the call setup of the CS fallback procedure. Details of PS connection switching can be found in [1].

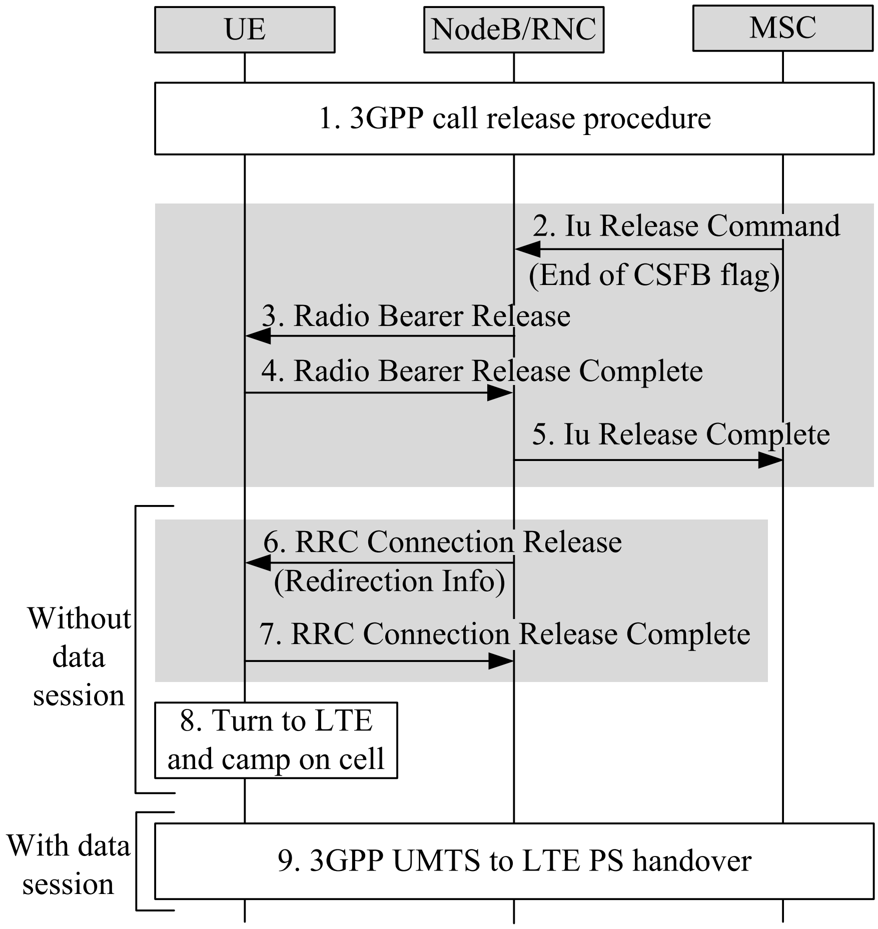


Figure 3: Call Release with IR

2.2 Call Release with Immediate-Return

Figure 3 illustrates the call release procedure with *Immediate-Return* (IR). After a voice call is released, if no UMTS data session is in progress, the UTRAN moves the UE to the LTE network immediately with the following steps:

Step 1. The standard 3GPP call release procedure is executed [8].

Steps 2-5. The MSC sends the UTRAN the Iu Release Command message to release the bearer between the MSC and the RNC. This message contains the *End of CS Fallback* (CSFB) flag to indicate that the call which was released is a CS fallback call. Then the NodeB sends the UE the Radio Bearer Release message to release the radio bearer between the NodeB and the UE.

Steps 6-8. According to the End of CSFB flag, the NodeB knows that the UE is LTE capable. The NodeB sends the UE the RRC Connection Release with Redirection Info message to release the radio connection between the NodeB and the UE. Then the UE switches to the LTE network according the redirection information in the message.

If the UE is engaged in a data session when the voice call is released, then Steps 6-8 are replaced by the standard 3GPP UMTS to LTE PS handover (Step 9) [1], and the data session is moved to the LTE network.

2.3 Call Release with Delayed-Return

When a voice call is released, if the UE is engaged in a data session, then it is switched back to LTE as shown in Figure 3 (Steps 1-5 and 9). If the UE is not engaged in a data session, then it does not need to return to LTE immediately. Figure 4 illustrates the call release procedure with *Delayed-Return* (DR). The UE releases the radio connection and stays in UMTS in the idle mode.

Steps 1-5 of the message flow in Figure 4 is the same as the call release procedure with IR. At Step 6, the NodeB sends the RRC Connection Release without Redirection Info message. Because this message does not contain the optional Redirection Info, the UE will not switch to LTE. This message instructs the UE to release the RRC connection, stay in UMTS, and change its status to the idle mode. Compared with the CS fallback with IR, Steps 8 and 9 in Figure 3 are saved in the CS fallback with DR.

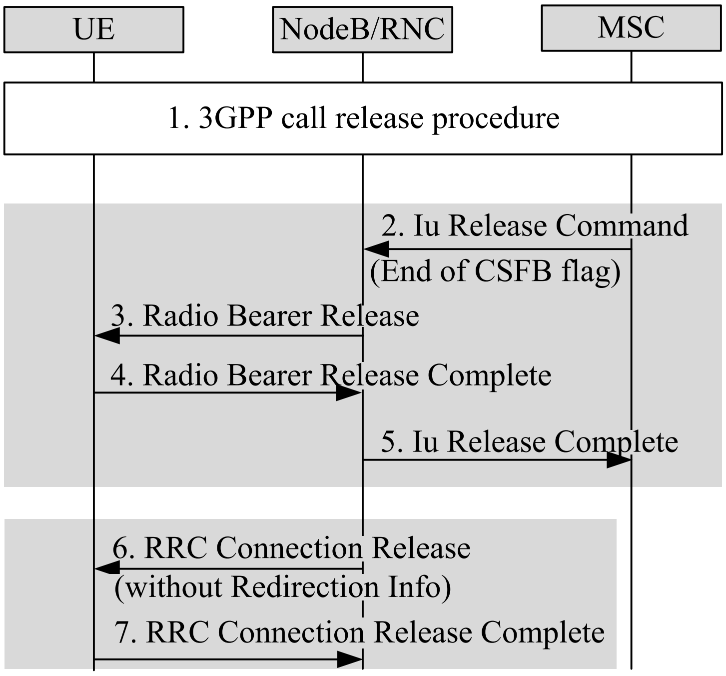


Figure 4: Call Release with DR

2.4 Data Session Setup in UMTS with Delayed-Return

Suppose that DR is applied, and the UE does not return to LTE after a voice call (i.e., there is no data session in progress when the voice call is released). If the next event to the UE is a data session arrival, then it will receive the PS paging message from the UMTS NodeB. The UE is switched to LTE to establish the PS connection. The detailed steps are described as follows (see Figure 5):

Step 1. The UE executes the 3GPP Inter-*Radio Access Technology* (RAT) cell reselection procedure from UTRAN [9] to perform the measurement process, and then selects a LTE cell.

Steps 2-4. The UE exchanges with the eNodeB the RRC Connection Request and Setup message pair to establish the radio connection. Then the UE sends the eNodeB the RRC Connection Setup Complete message to acknowledge the RRC connection establishment

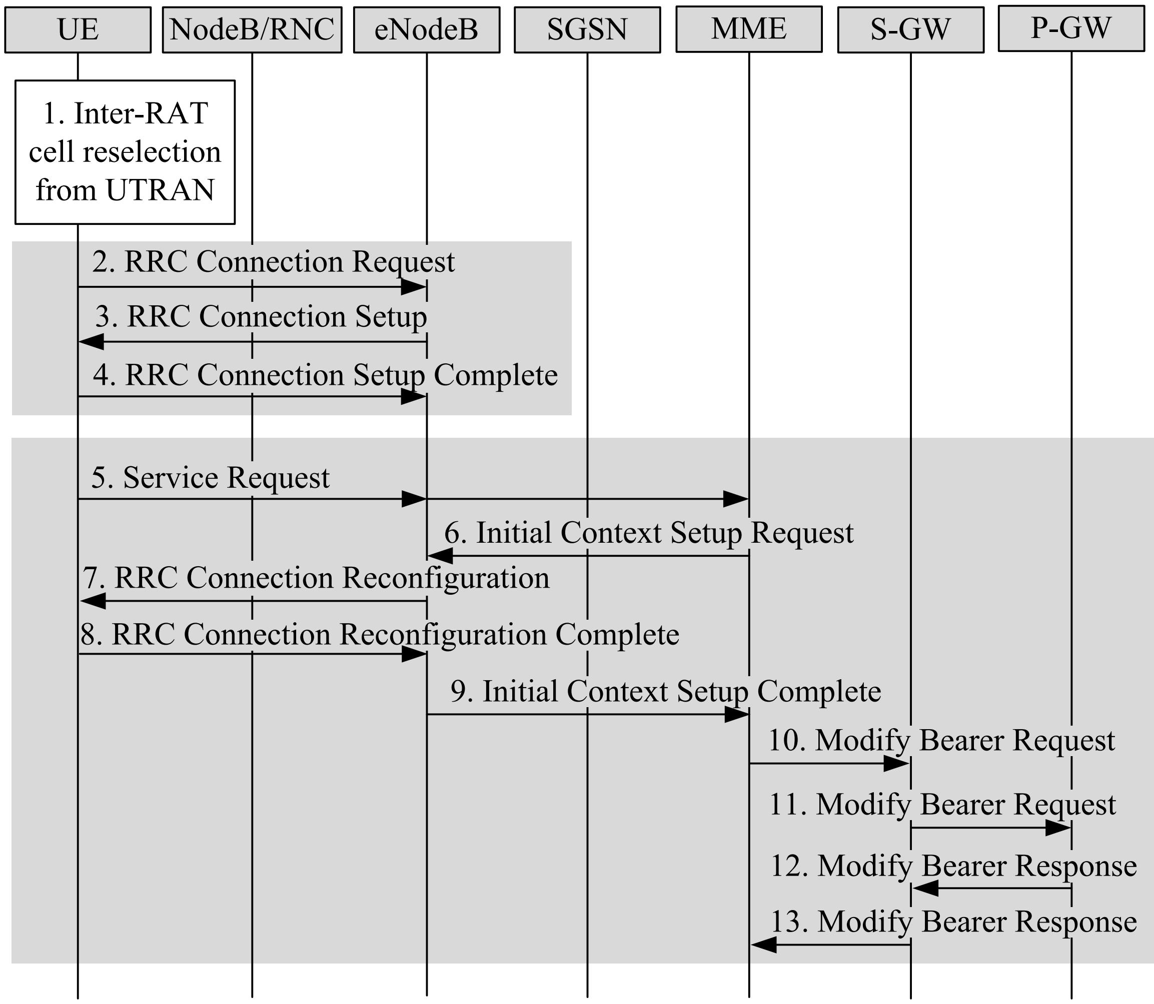


Figure 5: Date Session Setup in UMTS with DR

procedure.

Step 5. The UE sends the MME the Service Request message to initial the establishment of the bearer for the PS connection.

Steps 6-9. The MME exchanges with the eNodeB the Initial Context Setup Request and Complete message to establish the radio bearer between the UE and the eNodeB. The Request message also contains the bearer information between the eNodeB and the S-GW. The eNodeB sends the UE the RRC Connection Reconfiguration message to modify the bearer information of the radio connection. Then the UE sends the eNodeB the

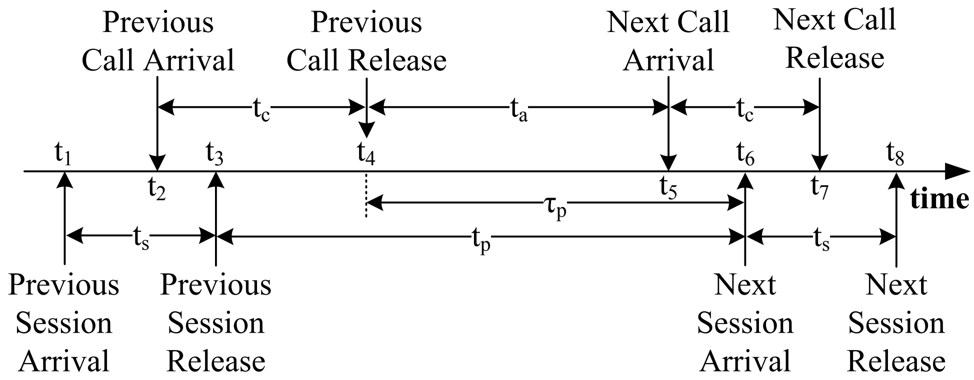


Figure 6: Timing Diagram for Voice Call and Data Session Arrivals

RRC Connection Reconfiguration Complete message to acknowledge the radio bearer reconfiguration.

Steps 10-13. The MME sends the S-GW the Modify Bearer Request message to establish the bearer between the eNodeB and the S-GW and the bearer between the S-GW and P-GW.

We note that for data session setup, the CS fallback with DR does not incur extra overhead over IR from the network viewpoint. Specifically, Steps 10-13 in Figure 2 are executed by IR, which are the same as Steps 1-4 in Figure 5. IR also executes Steps 5-13 in Figure 5 when a data session arrives.

3 Analytic Model

This section proposes an analytic model to study the performance improvement of the DR scheme over the IR scheme. Specifically, we derive the probability p that when a voice call arrives, the UE can be connected at UMTS without CS fallback due to DR. Figure 6 illustrates a timing diagram for voice call arrivals (at t_2 and t_5) and data session arrivals (at t_1 and t_6). Let $t_c = t_4 - t_2$ (also $t_7 - t_5$) be a voice call holding time. Let the inter-call arrival time $t_a = t_5 - t_4$ be a random variable with the density function $f_a(\cdot)$, the distribution function $F_a(\cdot)$, the variance V_a and the Laplace transform $f_a^*(s)$. Let the session holding time $t_s = t_3 - t_1$ (also $t_8 - t_6$) be a random variable with the mean $1/\mu$, and the inter-session arrival time $t_p = t_6 - t_3$ be a random variable with the density function $f_p(\cdot)$, the variance V_p and the Laplace transform $f_p^*(s)$. Suppose that the call release event at t_4 is a random observer of the period $[t_3, t_6]$. From the residual life theorem [10], the interval $\tau_p = t_6 - t_4$ is the residual life of t_p with the density function $r_p(\cdot)$, the distribution function $R_p(\cdot)$, and the Laplace transform $r_p^*(s)$. We define an *observation interval* as a period between when the previous call arrives and when the next call arrives (e.g., the interval $[t_2, t_5]$ in Figure 6). It is clear that the probability p described at the beginning of this section is the probability that no data session is in progress when the previous call is released (with probability p_1) and no data session arrives before the next voice call arrives (with probability p_2).

The sequence of t_s and t_p forms an alternating renewal process [11], and therefore $p_1 = \frac{E[t_p]}{E[t_p]+E[t_s]}$. Since a call release event is a random observer of t_s and t_p, p_2 can be expressed as $\Pr[t_a < \tau_p]$. According to the above description, we have

$$p = p_1 p_2 = \left(\frac{E[t_p]}{E[t_p] + E[t_s]}\right) \Pr[t_a < \tau_p] \tag{1}$$

Based on the inverse Laplace transform formula and the residue theorem [12, 13], $\Pr[t_a < \tau_p]$ in (1) is derived as

$$\begin{aligned}
\Pr[t_a < \tau_p] &= \int_{\tau_p=0}^{\infty} r_p(\tau_p) \int_{t_a=0}^{\tau_p} f_a(t_a) dt_a d\tau_p \\
&= \int_{\tau_p=0}^{\infty} r_p(\tau_p) F_a(\tau_p) d\tau_p \\
&= \left(\frac{1}{2\pi i}\right) \int_{\sigma-i\infty}^{\sigma+i\infty} \int_{\tau_p=0}^{\infty} r_p(\tau_p) e^{s\tau_p} \left[\frac{f_a^*(s)}{s}\right] d\tau_p ds \\
&= \left(\frac{1}{2\pi i}\right) \int_{\sigma-i\infty}^{\sigma+i\infty} r_p^*(-s) \left[\frac{f_a^*(s)}{s}\right] ds \\
&= -\sum_{z\in\sigma_p} Res_{s=z} r_p^*(-s) \left[\frac{f_a^*(s)}{s}\right]
\end{aligned} \tag{2}$$

where $i = \sqrt{-1}$, σ is a sufficiently small positive number, σ_p is the set of poles of $r_p^*(-s)$ in the right half of the complex plane, and $Res_{s=z}$ denotes the residue at the pole $s = z$. Alternatively, $\Pr[t_a < \tau_p]$ can also be derived as

$$\begin{aligned}
\Pr[t_a < \tau_p] &= \int_{t_a=0}^{\infty} f_a(t_a) \int_{\tau_p=t_a}^{\infty} r_p(\tau_p) d\tau_p dt_a \\
&= \int_{t_a=0}^{\infty} f_a(t_a)[1 - R_p(t)] dt_a \\
&= 1 - \left(\frac{1}{2\pi i}\right) \int_{\sigma-i\infty}^{\sigma+i\infty} \int_{t_a=0}^{\infty} f_a(t_a) e^{st_a} \left[\frac{r_p^*(s)}{s}\right] dt_a ds \\
&= 1 + \sum_{z\in\sigma_a} Res_{s=z} f_a^*(-s) \left[\frac{r_p^*(s)}{s}\right] \qquad (3)
\end{aligned}$$

where σ_a is the set of poles of $f_a^*(-s)$ in the right half of the complex plane.

For the demonstration purpose, we compute $\Pr[t_a < \tau_p]$ based on two cases of t_a and t_p distributions.

Case 1. t_a is a Gamma random variable with the shape parameter k and the rate parameter λ, and t_p is an Erlang random variable with the shape parameter m and the rate parameter γ. In this case, $k > 0$ is a real number and $m \geq 1$ is an integer.

Case 2. t_a is an Erlang random variable with the shape parameter k and the rate parameter λ, and t_p is a Gamma random variable with the shape parameter m and the rate parameter γ. In this case, $k \geq 1$ is an integer and $m > 0$ is a real number.

The Gamma distribution is considered because this distribution is widely used in telecom modeling [14]-[17]. We also select the Erlang distribution because this distribution can be easily extended into a hyper-Erlang distribution, which has been proven to be a good approximation to many other distributions as well as measured data [13, 18]. The Laplace transforms for the t_a and the t_p are

$$f_a^*(s) = \left(\frac{\lambda}{s+\lambda}\right)^k \text{ and } f_p^*(s) = \left(\frac{\gamma}{s+\gamma}\right)^m \tag{4}$$

For Gamma t_a and t_p, k and $m > 0$ are positive real numbers in (4). For Erlang t_a and t_p, k and m are positive integer numbers. From the residual life theorem [10] and (4), $r_p^*(s)$ is expressed as

$$r_p^*(s) = \left(\frac{\gamma}{sm}\right)[1 - f_p^*(s)] = \left(\frac{\gamma}{sm}\right)\left[1 - \left(\frac{\gamma}{s+\gamma}\right)^m\right] \tag{5}$$

For case 1, we substitute (4) and (5) into (2) to yield

$$\Pr[t_a < \tau_p] = \sum_{i=0}^{m-1}\sum_{j=0}^{i}\left[\frac{\gamma^{i-j}}{m\lambda^{i-j}(i-j)!}\right]\left(\frac{\lambda}{\gamma+\lambda}\right)^{k+i-j}\prod_{l=1}^{i-j}(k+l-1) \tag{6}$$

Note that in (6), when $i - j = 0$, $\prod_{l=1}^{i-j}$ represents an empty product, and its value is 1. From (1) and (6), p is re-written as

$$p = \sum_{i=0}^{m-1}\sum_{j=0}^{i}\left[\frac{\mu\gamma^{i-j}}{(\gamma+m\mu)\lambda^{i-j}(i-j)!}\right]\left(\frac{\lambda}{\gamma+\lambda}\right)^{k+i-j}\prod_{l=1}^{i-j}(k+l-1) \tag{7}$$

On the other hand, for case 2, from (4) and (5), (3) is re-written as

$$\Pr[t_a < \tau_p] = 1 - \sum_{i=0}^{k-1}\left\{\frac{\gamma}{\lambda m} - \sum_{j=0}^{i}\left(\frac{\lambda^{j-1}}{j!m\gamma^{j-1}}\right)\left(\frac{\gamma}{\lambda+\gamma}\right)^{m+j}\prod_{l=1}^{j}(m+l-1)\right\} \tag{8}$$

From (1) and (8), p is re-written as

$$p = \left(\frac{m\mu}{\gamma+m\mu}\right)\left\{1 - \sum_{i=0}^{k-1}\left\{\frac{\gamma}{\lambda m} - \sum_{j=0}^{i}\left(\frac{\lambda^{j-1}}{j!m\gamma^{j-1}}\right)\left(\frac{\gamma}{\lambda+\gamma}\right)^{m+j}\prod_{l=1}^{j}(m+l-1)\right\}\right\} \tag{9}$$

Equations (7) and (9) are validated against the discrete event simulation experiments, which shows that the discrepancies between the analytic and simulation results are within 0.5%.

4 Numerical Examples

This section studies the call setup delays of DR and IR. Let t_f be the time that the UE falls back from LTE to UMTS (i.e., Steps 1-10 in Figure 2). Let t_d be the UMTS outgoing call setup delay without the CS fallback (i.e., Steps 11-15 in Figure 2). Then the performance improvement α of the DR scheme over the IR scheme can be defined as

$$\alpha = 1 - \frac{(1-p)E[t_f] + E[t_d]}{E[t_f] + E[t_d]} = \frac{pE[t_f]}{E[t_f] + E[t_d]} \tag{10}$$

In (10), $E[t_f]+E[t_d]$ is the expected total call setup delay for IR, and $(1-p)E[t_f]+E[t_d]$ is the expected total call setup delay for DR. The larger the α value, the better the performance of DR over IR. From the call setup delay measurement of Qualcomm (see Section 2), $E[t_f] = 2.5$ seconds for 3GPP R8, $E[t_f] = 0.5$ seconds for 3GPP R9, and $E[t_d] = 4$ seconds. From Huawei's measurements [6], $E[t_f] = 9$ seconds for 3GPP R8, $E[t_f] = 3$ seconds for 3GPP R9, and $E[t_d] = 5$ seconds. We also measured the call setup delay at Broadband Mobile Lab of National Chiao Tung University [7], where $E[t_f]$ is more than 10 seconds and $E[t_d] = 7$ seconds. Our measurement results are more consistent with Huawei's results than that of Qualcomm's results. In this paper, we use Huawei's results to compute α in (10).

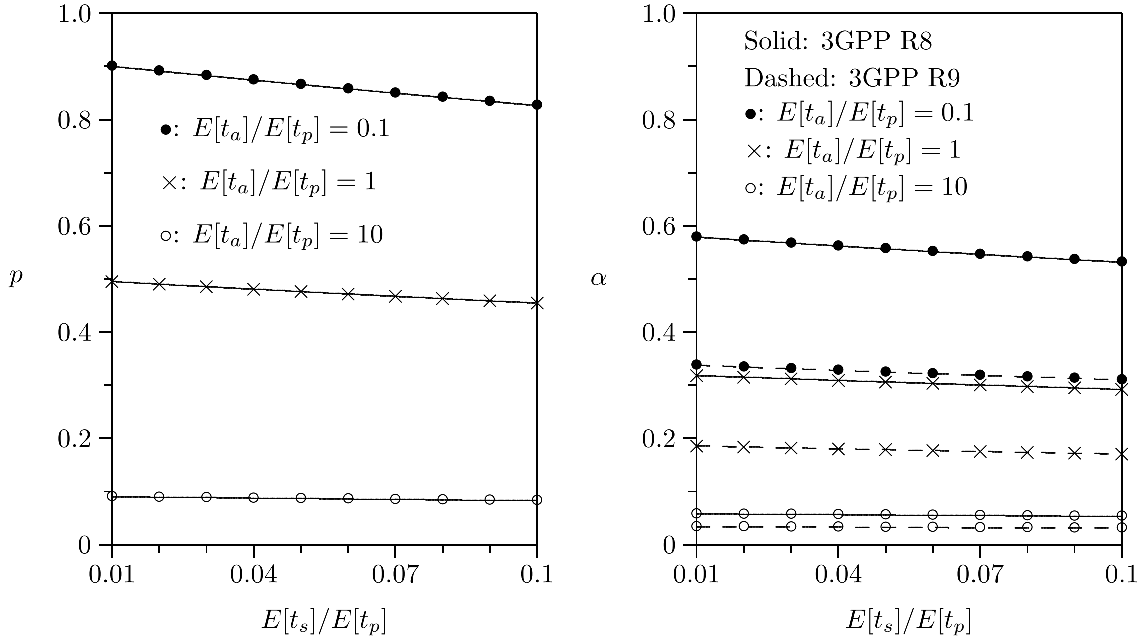
We also note that although an LTE data connection is "always on", the connection is in the idle mode (and is actually disconnected) if no data session is in progress. Because the expected session holding time is typically shorter than the expected inter-session arrival time [19, 20], we assume that $0.01E[t_p] \leq E[t_s] \leq 0.1E[t_p]$. We consider the effects of t_s (the session holding time), t_a (the inter-call arrival time), and t_p (the inter-session arrival time) on the probability p that a voice call can be connected without the CS fallback overhead. Note that the voice call holding time t_c does not affect p and is not considered. We also note that the α value is proportional to the p value (see (10)), and the effects on α are similar to those on p. Finally, to simplify our discussion, t_s and t_a are normalized by t_p.

Effects of $E[t_s]/E[t_p]$: Figure 7 (a) shows that p decreases as $E[t_s]/E[t_p]$ increases. When $E[t_s]/E[t_p]$ increases, a call is more likely to be released in the t_s interval. In this case, the UE will return to LTE immediately, and smaller p is observed. The non-trivial observation is that $E[t_s]/E[t_p]$ has insignificant impact on p for all $E[t_s]/E[t_p]$ values under our study. The probability p and the improvement α decrease by 8% when $E[t_s]$ increases from $0.01E[t_p]$ to $0.1E[t_p]$. In other words, in this operational range, we can ignore the effect of $E[t_s]/E[t_p]$ and can focus more on other parameters.

Effects of $E[t_a]/E[t_p]$: Figures 7 (a), 8 (a), and 9 (a) indicate that p decreases as $E[t_a]/E[t_p]$ increases. When $E[t_a]/E[t_p]$ increases, the data session is more likely to arrive before the voice call arrives (i.e., the UE will return to LTE before the next call arrives). Thus, a smaller p is observed. Figures 8 (a) and 9 (a) show that the effects of $E[t_a]/E[t_p]$ become insignificant when V_a or V_p is large, where large p and α are always observed.

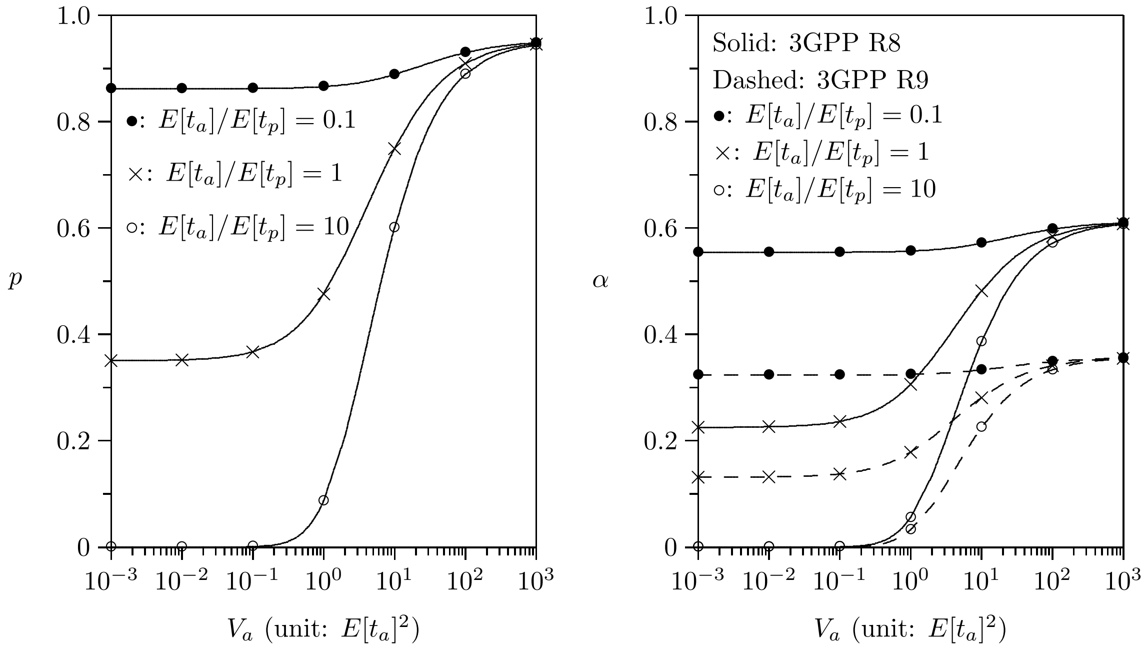
Effects of V_a: Figure 8 (a) indicates that p increases as V_a increases. For a fixed $E[t_a]$ value, when V_a increases, there are much more short t_a intervals than long t_a intervals. For short t_a, it is very likely that $t_a < \tau_p$ (i.e., larger $\Pr[t_a < \tau_p]$ is observed). From (1) and (10), p and α increase as V_a increases.

Effects of V_p: Figure 9 (a) shows that p increases as V_p increases. When the inter-session arrival interval becomes more irregular (i.e., V_p increases), more long and short t_p intervals are observed. Since the call release events are more likely to fall in long t_p intervals and the next calls are likely to arrive before the next sessions arrive, larger p



(a) Effects on p (b) Effects on α

Figure 7: Effects of $E[t_s]/E[t_p]$ and $E[t_a]/E[t_p]$ on p and α ($V_a = E[t_a]^2$ and $V_p = E[t_p]^2$)



(a) Effects on p (b) Effects on α

Figure 8: Effects of V_a and $E[t_a]/E[t_p]$ on p and α ($E[t_s]/E[t_p] = 0.05$ and $V_p = E[t_p]^2$)

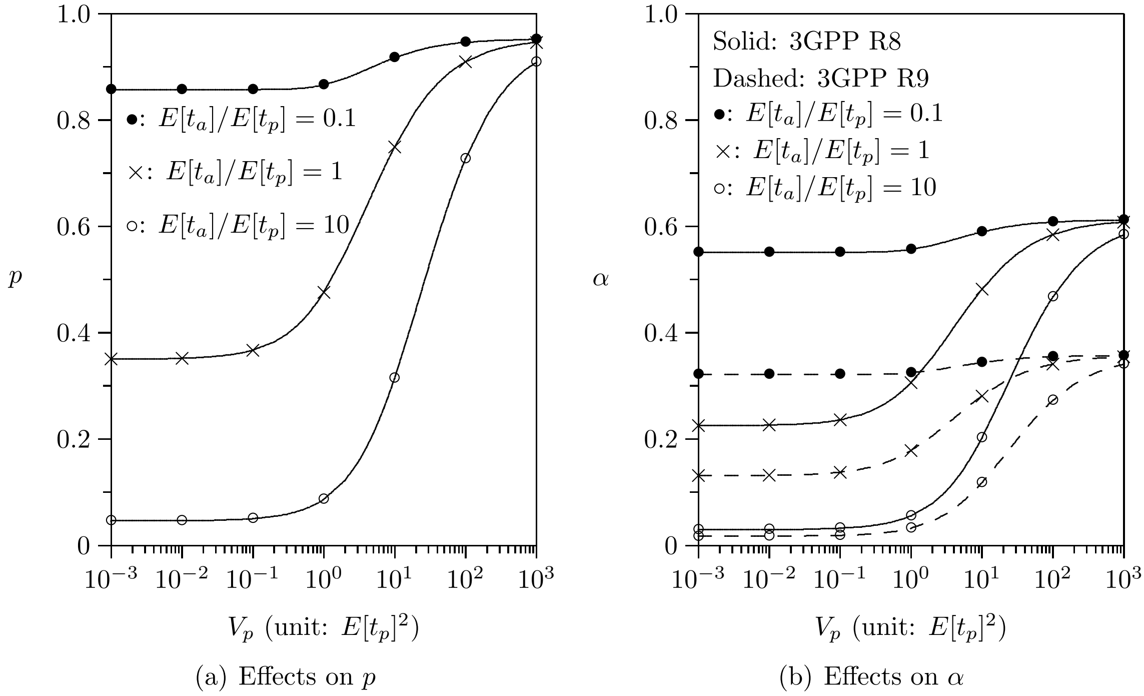


Figure 9: Effects of V_p and $E[t_a]/E[t_p]$ on p and α ($E[t_s]/E[t_p] = 0.05$ and $V_a = E[t_a]^2$)

and α are observed.

Based on Figures 7 (a), 8 (a), 9 (a), and Equation (10), Figures 7 (b), 8 (b), and 9 (b) plot the α curves against $E[t_s]/E[t_p]$, V_a, and V_p. These figures show that with probability p, the CS fallback with DR can reduce up to 60% outgoing call setup delay over the CS fallback with IR. We note that the DR scheme can also reduce the incoming call setup delay (i.e., the delay between when the network pages the UE and when the UE rings). The incoming call setup delay is typically shorter than the outgoing call setup delay. From (10), the DR scheme has even better α performance for the incoming calls than that for the outgoing calls. Since the called party of a voice call does not experience call setup delay, the improvement α is only meaningful from the network cost viewpoint.

5 Conclusions

This paper proposed the DR scheme to avoid unnecessary CS fallbacks. Analytic model was developed based on the real LTE/UMTS network measurements to compare the DR scheme with the existing IR scheme. The performance is measured by the probability p that when a voice call arrives, the UE can be connected at UMTS without CS fallback, and therefore, non-necessary switching between UMTS and LTE is avoided. In other words, when a voice call arrives, the UE does not need to switch from LTE to UMTS, and when the call is complete, the UE does not need to switch from UMTS to LTE. Our study indicated that the DR scheme can effectively improve the CS fallback performance when

- the inter-call arrival time t_a is short (i.e., the voice calls arrive frequently),
- the variance of t_a is large (i.e., the inter-call arrival time is irregular), or
- the variance of the inter-session arrival time t_p is large (i.e., the inter-session arrival time is irregular).

The last two items of our conclusions are not trivial, and are used as guidelines to further investigate the user behavior by a commercial mobile operator. For users with long inter-call arrival time and regular call and data session arrivals, the CS fallback with IR is exercised, while for the users with short inter-call arrival time and irregular call and data session arrivals, the CS fallback with DR is exercised.

As a final remark, the DR scheme can be practically implemented in NodeB with a minor modification in the RRC Connection Release message. Therefore, the DR scheme is an effective approach for reducing the CS fallback costs. In the future, we will investigate the DR scheme by the call and data traffic statistics collected from the commercial mobile telecom network. We will also consider other approaches to avoid unnecessary CS fallbacks (e.g., a timer-based scheme that determines the optimal time interval for the UE to stay in

UMTS based on different traffic rates). Moreover, because both CS fallback and *Enhanced Single Radio Voice Call Continuity* (eSRVCC) [21, 22] are voice call solutions in LTE, we will compare the call performance between these two solutions.

References

[1] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 12), Technical Specification 3GPP TS 23.401 V12.2.0*, 2013.

[2] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS); Service description; Stage 2 (Release 12), Technical Specification 3GPP TS 23.060 V12.2.0*, 2013.

[3] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Core Network; Digital cellular telecommunications system (Phase 2+); Basic call handling; Technical realization (Release 1998), Technical Specification 3GPP TS 03.18 V7.5.0*, 2003.

[4] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Circuit Switched (CS) fallback in Evolved Packet System (EPS); Stage 2 (Release 11), Technical Specification 3GPP TS 23.272 V12.0.0*, 2013.

[5] Circuit-switched fallback. The first phase of voice evolution for mobile LTE devices. Qualcomm Incorporated. USA. [online]. Available: http://www.qualcomm.com/media/documents/files/circuit-switched-fallback-the-first-phase-of-voice-evolution-for-mobile-lte-devices.pdf

[6] Circuit-Switched Fallback, Ultra-Flash CSFB. Huawei Technologies Co. Ltd. China. [online]. Available: http://www.huawei.com/ilink/en/download/HW_267762

[7] Y.-B. Lin, P.-J. Lin, Y. C. Sung, Y.-K. Chen, W.-E. Chen, N. Alrajeh, B.-S. P. Lin, and C.-H. Gan, “Performance measurements of TD-LTE, WiMAX and 3G systems,” *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 153-160, Jun. 2013.

[8] Y.-B. Lin and A.-C. Pang, *Wireless and Mobile All-IP Networks*, New York: John Wiley & Sons, Inc., 2005.

[9] 3GPP, *3rd Generation Partnership Project Technical Specification Group Radio Access Network; Radio Resource Control (RRC); Protocol specification (Release 11), Technical Specification 3GPP TS 25.331 V11.7.0*, 2013.

[10] L. Kleinrock, *Queueing Systems, Vol. I, Theory*, New York: John Wiley & Sons, Inc., 1976.

[11] S. M. Ross, *Stochastic Processes*, New York: John Wiley & Sons, Inc., 1996.

[12] W. R. LePage, *Complex Variables and the Laplace Transform for Engineers*, New York: Dover, 1980.

[13] Y. Fang and I. Chlamtac, “Teletraffic analysis and mobility modeling of PCS networks,” *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1062-1072, Jul. 1999.

[14] I. F. Akyildiz, J. S. M. Ho, and Y.-B. Lin, “Movement-based location update and selective paging for PCS networks,” *IEEE/ACM Trans. Netw.*, vol. 4, no. 4, pp. 629-638, Aug. 1996.

[15] D.-W. Huang, P. Lin, and C.-H. Gan, “Design and performance study for a mobility management mechanism (WMM) using location cache for wireless mesh network,” *IEEE Trans. Mobile Comput.*, vol. 7, no. 5, pp. 546-556, May 2008.

[16] C.-Y. Hong and A.-C. Pang, “3-approximation algorithm for joint routing and link scheduling in wireless relay networks,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 856-861, Feb. 2009.

[17] R.-H. Liou, Y.-B. Lin, and S.-C. Tsai, "An investigation on LTE mobility management," *IEEE Trans. Mobile Comput.*, vol. 12, no. 2, pp. 166-176, Jan. 2013.

[18] F. Kelly, *Reversibility and Stochastic Networks*, New York: John Wiley & Sons, Inc., 1979.

[19] J. Färber, S. Bodamer, J. Charzinski, "Measurement and modelling of internet traffic at access networks," in *Proc. EUropean Network of universities and companies in Inform. and Commun. Eng. (EUNICE)*, Munich, Germany, 1998, pp. 196-203.

[20] J. Ferreira and F. J. Velez, "Enhanced UMTS services and applications characterization," *Telektronikk*, vol. 101, no. 1, pp. 113-131, 2005.

[21] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; IP Multimedia (IM) Core Network (CN) subsystem IP Multimedia Subsystem (IMS) service continuity; Stage 3, Technical Specification 3GPP TS 24.237 V12.2.0*, 2013.

[22] Y.-B. Lin, R.-H. Liou, Y. C. Sung, and P.-C. Cheng, "Performance evaluation of LTE eSRVCC with limited access transfers," submitted to *IEEE Trans. Wireless Commun.*