

A New Approach for Serving Radio Network Controller Relocation in UMTS All-IP Network

Ai-Chun Pang*, Yi-Bing Lin[†], Hsien-Ming Tsai[‡] and Prathima Agrawal[§]

*Dept. of Comp. Sci. & Info. Engr., National Taiwan University

Taipei, Taiwan, R.O.C.; Email: acpang@csie.ntu.edu.tw

[†]Chair Professor, Providence University

Taichung, Taiwan, R.O.C.; Email: liny@csie.nctu.edu.tw

[‡]Quanta Research Inst., Quanta Computer Inc.

Taoyuan, Taiwan, R.O.C.; Email: samuel.tsai@quantatw.com

[§]Internet Architecture Research Lab., Telcordia Technologies

Morristown, NJ 07960-6438; Email: pagrawal@research.telcordia.com

Abstract—To support real-time multimedia services in UMTS all-IP network, 3GPP TR 25.936 proposed two approaches to support real-time *Serving Radio Network Controller* (SRNC) switching, which require packet duplication during SRNC relocation. These approaches significantly consume extra system resources. This paper proposes the *fast SRNC relocation* (FSR) approach that does not duplicate packets. In FSR, a packet buffering mechanism is implemented to avoid packet loss at the target RNC. We propose an analytic model to investigate the performance of FSR. The numerical results show that packet loss at the source RNC can be ignored. Furthermore, the expected number of packets buffered at the target RNC is small, which does not prolong packet delay.

I. INTRODUCTION

Mobility, privacy and immediacy offered by wireless access commonly create new opportunities for Internet business, and mobile networks are becoming a platform that provides leading-edge Internet services. Through integration of the Internet and the third generation (3G) wireless communication, next generation telecommunications networks will provide global information access for mobile users [1]. 3GPP [1], [5], [6] proposed the *Universal Mobile Telecommunications System* (UMTS) all-IP architecture to integrate the IP and wireless technologies, which has evolved from the GSM, *General Packet Radio Service* (GPRS), and UMTS Release 1999.

Figure 1 shows a UMTS all-IP network architecture (another UMTS all-IP option can be found in [1], [5]). In this figure, the dashed lines represent signaling links, and the solid lines represent data and signaling links. The UMTS all-IP network connects to the *Packet Data Network* (PDN; see Figure 1 (a)) or the *IP Multimedia Core Network Subsystem* (see Figure 1 (b)) through the *Serving GPRS Support Node* (SGSN; see Figure 1 (c)) and the *Gateway GPRS Support Node* (GGSN; see Figure 1 (d)). The SGSN connects to the radio access network. The GGSN provides interworking with the external PDN, and is connected with SGSNs via an IP-based GPRS

backbone network. Both the GGSN and SGSN communicate with the *Home Subscriber Server* (see Figure 1 (e)) to obtain mobility and session management information of subscribers. The *UMTS Terrestrial Radio Access Network* (UTRAN) consists of *Node Bs* (the UMTS term for base stations; see Figure 1 (f)) and *Radio Network Controllers* (RNCs; see Figure 1 (g)) connected by an ATM network. A *user equipment* (UE; see Figure 1 (h)) communicates with one or more Node Bs through the radio interface *Uu* based on the *Wideband CDMA* (WCDMA) radio technology [8].

In the UMTS all-IP network, the IP packets are routed between the UE and the GGSN. By using the *Packet Data Protocol* (PDP) context activation procedure [4], a PDP context is created to establish the routing path for IP packet delivery. Besides the packet routing information (e.g., the UE's IP address), the PDP context also contains the QoS profiles and other parameters. Due to the CDMA characteristics, multiple radio paths (for delivering the same IP packets) may exist between the UE and more than one Node Bs. An example of multiple routing paths is illustrated in Figure 2 (a). In this figure, an IP-based *GPRS Tunneling Protocol* (GTP) connection is established between the GGSN and RNC1. The UE connects to two Node Bs (B1 and B2). Node B1 is connected to RNC1, and Node B2 is connected to RNC2. An Iur link between RNC1 and RNC2 is established so that the signal (i.e., IP packets) sent from the UE to Node B2 can be forwarded to RNC1 through RNC2. RNC1 then combines the signals from Node B1 and B2, and forwards them to SGSN1. Similarly, the packets sent from the GGSN to RNC1 will be forwarded to both Node B1 and RNC2 (and then Node B2). In this example, RNC1 is called the *Serving RNC* (SRNC). RNC2 is called the *Drift RNC* (DRNC), which transparently routes the packets through the Iub (between the Node B and the RNC) and Iur (between two RNCs) interfaces. Suppose that the UE moves from Node B1 toward Node B2, and the radio link between the UE and Node B1 is disconnected. In this case, the routing path will be <UE→Node B2→RNC2→RNC1→SGSN1→GGSN> as shown in Figure 2 (b).

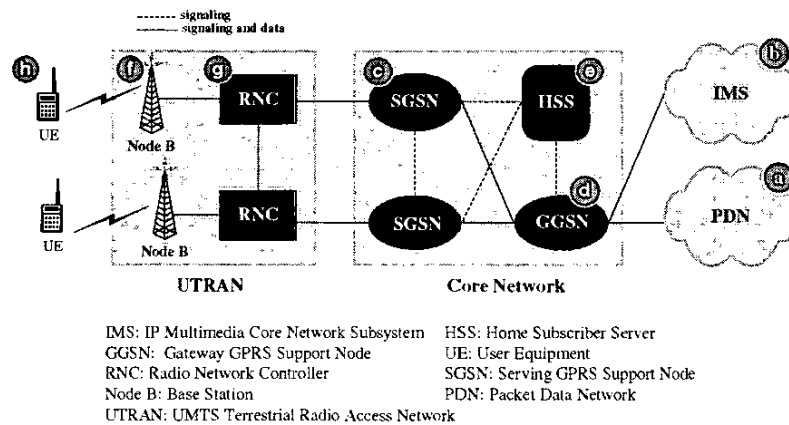


Fig. 1. The UMTS All-IP Network Architecture

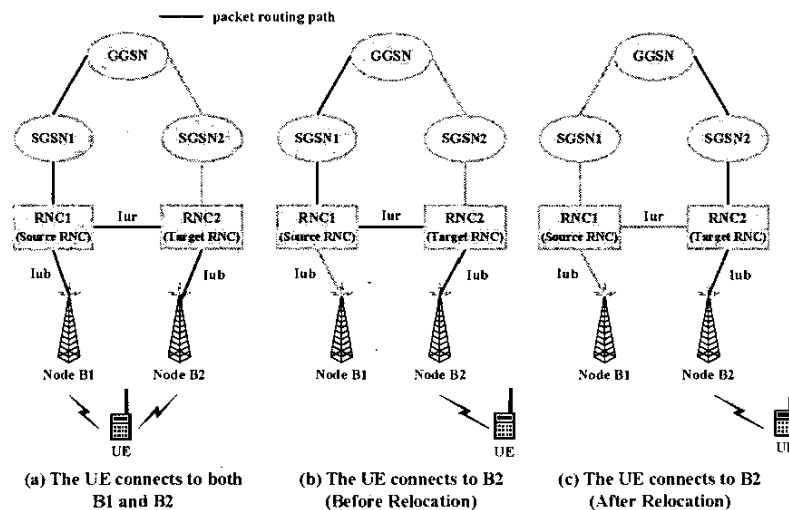


Fig. 2. SRNC Relocation

In this scenario, it does not make sense to route packets between the UE and the core network through RNC1. Therefore SRNC relocation may be performed to remove RNC1 from the routing path. After SRNC relocation, the packets are routed to the GGSN directly through RNC2 and SGSN2 (see Figure 2 (c)), and RNC2 becomes the SRNC.

In 3GPP TS 23.060 [4], a lossless SRNC relocation procedure was proposed for non-real-time data services. In this approach, in the beginning of SRNC relocation, the source RNC (RNC1 in Figure 2 (b)) first stops transmitting downlink IP packets to the UE. Then it forwards the next packets to the target RNC (RNC2 in Figure 2 (b)) via a GTP tunnel between the two RNCs. The target RNC stores all IP packets forwarded from the source RNC. After taking over the SRNC role, the target RNC restarts the downlink data transmission to the UE. In this approach, no packet is lost during the SRNC switching period. Unfortunately, this approach does not support real-time data transmission because the IP data traffic will be suspended for a long time during SRNC switching. In order to support real-time multimedia services, 3GPP TR 25.936 [3] proposes *SRNC duplication* (SD) and *core network bi-casting* (CNB). These two approaches duplicate data packets during SRNC relocation, which may not efficiently utilize system resources. In this paper, we propose a new approach called *fast SRNC relocation* (FSR) to provide real-time SRNC switching without packet duplication.

II. RELATED WORK

This section describes the previously proposed SRNC relocation procedures for real-time multimedia services; that is, *SRNC duplication* (SD) and *core network bi-casting* (CNB) proposed in 3GPP TR 25.936 [3].

A. SRNC Duplication (SD)

Consider Figure 2 (b). Suppose that the UE is connected to the source RNC and SGSN1 before performing SRNC relocation. The target RNC is the drift RNC, which is connected to the source RNC via the Iur interface. After SRNC relocation, the SRNC role is moved from the source RNC to the target RNC, and the IP packets for the UE are directly routed through SGSN2 and the target RNC (see Figure 2 (c)). Figure 3 shows the four stages of the SD procedure. Stage I (Figure 3 (a)) initiates SRNC relocation. In this stage, the user IP packets are delivered through the old path <GGSN→SGSN1→source RNC→target RNC→UE>. The following steps are executed.

- 1-2. When the Node B of the source RNC no longer connects to the UE, the source RNC initiates SRNC relocation. Specifically, the source RNC sends a **Relocation_Required** message (including the ID of the target RNC) to SGSN1.

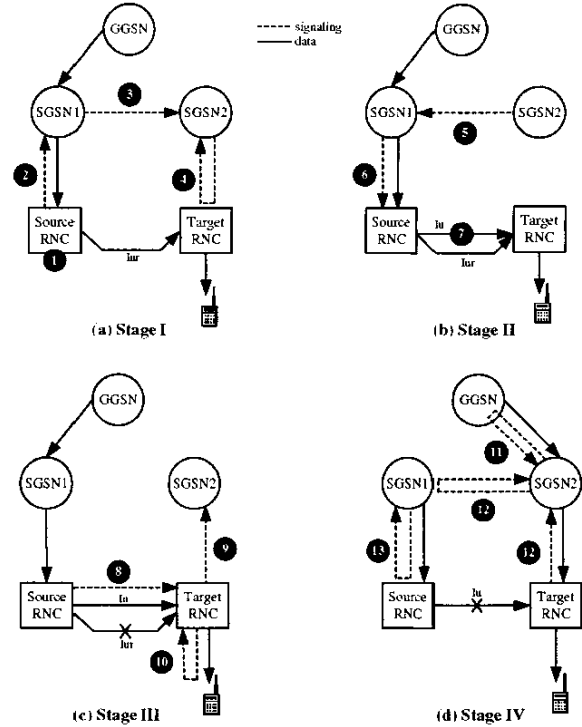


Fig. 3. The SRNC Duplication (SD) Approach

3. Based on the ID of the target RNC, SGSN1 determines if the SRNC relocation is intra-SGSN SRNC relocation or inter-SGSN SRNC relocation. Assume that it is inter-SGSN SRNC relocation. By sending a **Forward_Relocation_Request** message, SGSN1 requests SGSN2 to allocate the resources (to be described in Step 4) for the UE.
4. SGSN2 sends a **Relocation_Request** message with the **Radio Access Bearer** (RAB) parameters to the target RNC. The RAB parameters include the traffic class (e.g., conversational, streaming, interactive or background), traffic handling priority, maximum and guaranteed bit rates, and so on [2]. After all necessary resources for the RAB are successfully allocated, the target RNC sends a **Relocation_Request_Acknowledge** message to SGSN2.

In Stage II (Figure 3 (b)), a forwarding path <source RNC→target RNC→UE> for downlink packet delivery is created between the source and the target RNCs through the Iu interface. The source RNC duplicates the packets and forwards these packets to the target RNC. Thus the downlink packets are simultaneously transmitted through both the old path (via the Iur interface) and the forwarding path (via the Iu interface) between the source RNC and the target RNC. Note that 3G TR25.936 [3] did not clearly describe if an Iu link can be directly established between two RNCs. If not, an indirect path <source RNC→SGSN1→SGSN2→target RNC> is required.

To favor the SD approach, we assume a direct link between the source and target RNCs. The following steps are executed in Stage II.

- 5-6. SGSN2 sends a **Forward_Relocation_Response** message to SGSN1, which indicates that all resources (e.g., RAB) are allocated. SGSN1 forwards this information to the source RNC through a **Relocation_Command** message.
7. Upon receipt of the **Relocation_Command** message, the source RNC duplicates the downlink packets and transmits the duplicated packets to the target RNC through the forwarding path (via the Iu interface at the IP layer). The forwarded packets are discarded at the target RNC before it becomes the SRNC (i.e., before the target RNC receives the **Relocation_Commit** message at Step 8).

In Stage III (Figure 3 (c)), the Iur link between the source RNC and the target RNC (i.e., the old path) is disconnected. The downlink packets arriving at the source RNC are forwarded to the target RNC through the Iu link (i.e., the forwarding path). A data-forwarding timer is maintained in the source RNC. When the timer expires, the forwarding operation at the source RNC is stopped. The following steps are executed in Stage III.

8. With a **Relocation_Commit** message, the source RNC transfers *Serving Radio Network Subsystem* (SRNS) context (e.g., QoS profile for the RAB) to the target RNC.
9. Upon receipt of the **Relocation_Commit** message, the target RNC sends a **Relocation_Detect** message to SGSN2, which indicates that the target RNC will become the SRNC.
10. At the same time, the target RNC sends a **RAN_Mobility_Information** message to the UE. This message triggers the UE to send the uplink IP packets to the target RNC. After the UE has reconfigured itself, it replies the **RAN_Mobility_Information_Confirm** message to the target RNC.

In Stage IV (Figure 3 (d)), the packet routing path is switched from the old path to the new path $\langle \text{GGSN} \leftrightarrow \text{SGSN2} \leftrightarrow \text{target RNC} \leftrightarrow \text{UE} \rangle$. At this stage, the target RNC becomes the SRNC. The source RNC forwards the downlink packets to the target RNC until the data-forwarding timer expires. The following steps are executed in Stage IV.

11. SGSN2 sends a **Update_PDP_Context_Request** message to the GGSN. Based on the received message, the GGSN updates the corresponding PDP context and returns a **Update_PDP_Context_Response** message to SGSN2. Then the downlink packet routing path is switched from the old path to the new

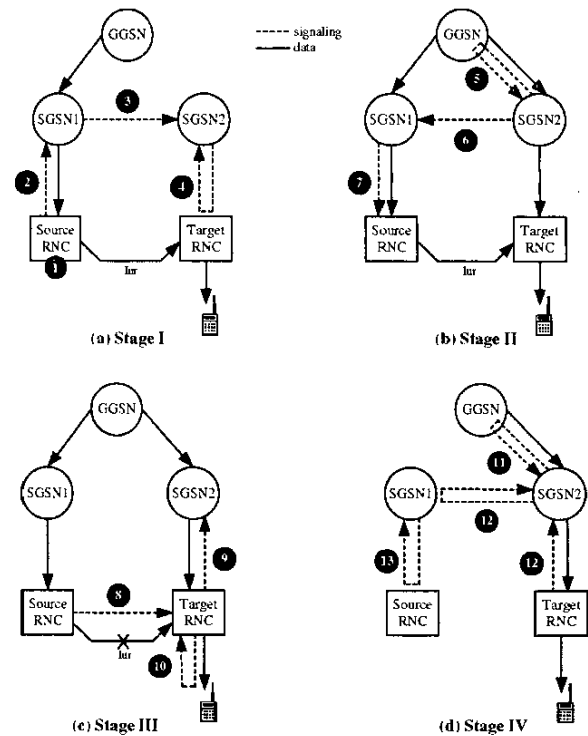


Fig. 4. The Core Network Bi-casting (CNB) Approach

path. At this moment, the target RNC receives the downlink packets from two paths (i.e., the forwarding and new paths), and transmits them to the UE. Since the transmission delays for these two paths are not the same, the packets arriving at the target RNC may not be in sequence, which results in out-of-order delivery.

12. By sending the **Relocation_Complete** message to SGSN2, the target RNC indicates the completion of the relocation procedure. Then SGSN2 exchanges this information with SGSN1 using the **Forward_Relocation_Complete** and **Forward_Relocation_Complete_Acknowledge** message pair.
13. Finally, SGSN1 sends an **Iu_Release_Command** message to request the source RNC to release the Iu connection in the forwarding path. When the data-forwarding timer expires, the source RNC replies an **Iu_Release_Complete** message.

B. Core Network Bi-casting (CNB)

Figure 4 shows the four stages of the CNB procedure when the communicating UE moves from the source RNC to the target RNC. Stage I (Steps 1-4, Figure 4 (a)) is the same as Stage I in SD, which requests the target RNC to allocate the necessary resources for relocation.

In Stage II (Figure 4 (b)), the downlink packets are duplicated at the GGSN, and are sent to the target RNC through both the old path <GGSN→SGSN1→source RNC→target RNC> and the new path <GGSN→SGSN2→target RNC>. The following steps are executed.

5. Upon receipt of the **Relocation.Request.Acknowledge** message at Step 4, SGSN2 sends a **Update.PDP.Context.Request** message that requests the GGSN to bi-cast the downlink packets. The GGSN starts to perform bi-casting and replies SGSN2 a message **Update.PDP.Context.Response**. At this moment, the downlink packets are simultaneously transmitted to the target RNC through the old and the new paths. Since the target RNC has not taken the SRNC role (i.e., the target RNC has not received the **Relocation.Commit** message), the packets routed through the new path are discarded at the target RNC.
- 6-7. These steps are used to inform the source RNC that all necessary resources are allocated, which are similar to Steps 5 and 6 in the SD approach.

In Stage III (Figure 4 (c)), the Iur link between the source RNC and the target RNC is disconnected, and the downlink packets arriving at the source RNC are discarded.

- 8-10. These steps are used to move the SRNC role from the source RNC to the target RNC, which are similar to Steps 8-10 in the SD approach.

In Stage IV (Figure 4 (d)), the GGSN is informed to stop downlink packet bi-casting. The target RNC takes the SRNC role to transmit the downlink packets to the UE.

11. Through the **Update.PDP.Context.Request** message, SGSN2 informs the GGSN to stop downlink packet bi-casting. Then the GGSN removes the GTP tunnel between the GGSN and SGSN1, and replies SGSN2 the **Update.PDP.Context.Response** message.
12. With the **Relocation.Complete** message, the target RNC informs SGSN2 that the relocation procedure is successfully performed. Then SGSN2 exchanges this information with SGSN1 using the **Forward.Relocation.Complete** and **Forward.Relocation.Complete.Acknowledge** message pair.
13. Finally, SGSN1 and the source RNC exchange the **Iu.Release.Command** and **Iu.Release.Complete** message pair to release the Iu connection in the old path.

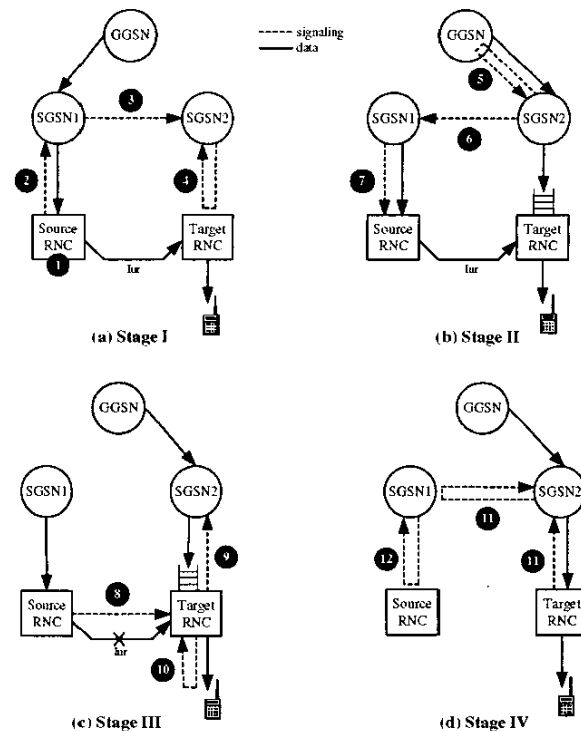


Fig. 5. The Fast SRNC Relocation (FSR) Approach

III. FAST SRNC RELOCATION (FSR)

This section describes the FSR approach and compares this approach with SD and CNB. As shown in Figure 2 (b), the UE is connected to the source RNC and SGSN1 before SRNC relocation. After relocation, the data packets for the UE are directly routed through the target RNC and SGSN2 as shown in Figure 2 (c). Figure 5 illustrates the four stages of the FSR procedure.

Stage I (Figure 5 (a)) initiates SRNC relocation. In this stage, the routing path of downlink packets is <GGSN→SGSN1→source RNC→target RNC→UE>. The following steps are executed in Stage I.

- 1-2. When the Node B of the source RNC no longer connects to the UE, the source RNC initiates SRNC relocation and sends the ID of the target RNC to SGSN1 through the **Relocation.Request** message.
3. Based on the ID of the target RNC, SGSN1 determines that it is inter-SGSN SRNC relocation. SGSN1 requests SGSN2 to allocate the resources for the UE through the **Forward.Relocation.Request** message.
4. SGSN2 and the target RNC exchange the **Relocation.Request** and **Relocation.Request.Acknowledge** message pair to allocate the necessary resources for the UE.

In Stage II (Figure 5 (b)), the GGSN routes the downlink packets to the old path before receiving the **Update_PDP.Context.Request** message (Step 5 in Figure 5 (b)). The packets delivered through the old path are called "old" packets. After the GGSN has received the **Update_PDP.Context.Request** message, the downlink packets are routed to the new path <GGSN→SGSN2→target RNC>. The packets delivered by the new path are called "new" packets. The "new" packets arriving at the target RNC are buffered until the target RNC takes over the SRNC role. The following steps are executed in Stage II.

5. Upon receipt of the **Relocation.Request.Acknowledge** message, SGSN2 sends a **Update_PDP.Context.Request** message to the GGSN. Based on the received message, the GGSN updates the corresponding PDP context fields and returns a **Update_PDP.Context.Response** message to SGSN2. Then the downlink packet routing path is switched from the old path to the new path. At this stage, the "new" downlink packets arriving at the target RNC are buffered.
- 6-7. SGSN2 sends a **Forward.Relocation.Response** message to SGSN1 to indicate that all resources for the UE are allocated. SGSN1 forwards this information to the source RNC through the **Relocation.Command** message.

In Stage III (Figure 5 (c)), the Iur link between the source RNC and the target RNC is disconnected. The "old" downlink packets arriving at the source RNC later than the **Relocation.Command** message (Step 7 in Figure 5 (b)) are dropped. In this stage, Steps 8-10 switch the SRNC role from the source RNC to the target RNC.

8. With the **Relocation.Commit** message, the SRNC context of the UE is transferred from the source RNC to the target RNC.
- 9-10. The target RNC sends a **Relocation.Detect** message to SGSN2. At the same time, the target RNC sends a **RAN.Mobility.Information** message to the UE, which triggers the UE to send the uplink IP packets through the new path <UE→target RNC→SGSN2→GGSN>.

By executing Steps 11 and 12 at Stage IV (Figure 5 (d)), the target RNC informs the source RNC that SRNC relocation is successfully performed. Then the source RNC releases the system resources for the UE.

11. The target RNC sends the **Relocation.Complete** message to SGSN2, which indicates that SRNC relocation is successfully performed. Then SGSN2 exchanges this information with SGSN1 through the **Forward.Relocation.Complete** and **Forward.Relocation.Complete.Acknowledge**

TABLE I
COMPARING FSR WITH SD AND CNB

Approaches Issues	FSR	SD	CNB
Packet Duplication	No	Yes	Yes
Packet Loss at Source RNC	Yes	Yes	No
Packet Loss at Target RNC	No	Yes	Yes
Packet Buffering	Yes	No	No
Out-of-order Delivery	No	Yes	No
Extra Signaling	No	No	Yes

message pair.

12. Finally, SGSN1 and the source RNC exchanges the **Iu.Release.Command** and **Iu.Release.Complete** message pair to release the Iu connection in the old path.

Based on the above discussions, Table I compares FSR with SD and CNB. The following issues are addressed.

Packet Duplication. During SRNC relocation, IP packets are duplicated at the source RNC in SD. Similarly, IP packets are duplicated at the GGSN in CNB. Packet duplication will significantly consume system resources. On the other hand, packet duplication is not needed in the FSR approach.

Packet Loss. Packet loss may occur in these three approaches either at the source RNC or at the target RNC. For SD and FSR, the data packets arriving at the source RNC may be lost. In SD, the "old" packets are dropped at the source RNC when the data-forwarding timer expires (Step 13 in Figure 3 (d)). In FSR, the "old" packets are dropped if they arrive at the source RNC later than the **Relocation.Command** message (see Step 7 in Figure 5 (b)) does.

For SD and CNB, the data packets may be lost at the target RNC. In SD, the target RNC discards the forwarded packets from the source RNC if these packets arrive at the target RNC earlier than the **Relocation.Commit** message does (Step 7 in Figure 3 (b)). In CNB, the duplicated packets may be lost at the target RNC because the packets from the new path are dropped before the target RNC becomes the SRNC (see Step 5 in Figure 4 (b)). On the other hand, since the packet buffering mechanism is implemented in FSR, the packets are not lost at the target RNC.

Packet Buffering. To avoid packet loss at the target RNC, the packet buffering mechanism is implemented in FSR, which is not found in both SD and CNB approaches.

Out-of-order Delivery. In SD, two paths (i.e., the forwarding and new paths) are utilized to simultaneously transmit the

downlink packets (see Step 11 in Figure 3 (d)). Since the transmission delays for these two paths are not the same, the packets arriving at the target RNC may not be in sequence, which results in out-of-order delivery. On the other hand, this problem does not exist in FSR and CNB because the target RNC in these two approaches only processes the packets from one path (either the old path or the new path) at any time, and the out-of-order packets are discarded (see Step 5 in Figure 4 (b)).

Extra Signaling. The SD approach follows the standard SRNC relocation procedure proposed in 3G 23.060 [4]. The FSR approach reorders the steps of the 3G 23.060 SRNC relocation procedure. Both approaches do not introduce any extra signaling cost. On the other hand, CNB exchanges additional *Update_PDP_Context_Request* and *Update_PDP_Context_Response* message pair (see Step 5 in Figure 4) between the GGSN and SGSN2, which incurs extra signaling cost. Note that all three approaches can be implemented in the GGSN, SGSN and RNC without introducing new message types to the existing 3GPP specifications.

In conclusion, SD and CNB require packet duplication that will double the network traffic load during SRNC relocation. For the SD approach, it is not clear if the Iu link in the forwarding path can be directly established between two RNCs. If not, an indirect path <source RNC→SGSN1→SGSN2→target RNC> is required. Also, it is not clear if the target RNC will be informed to stop receiving the forwarded packets when the data-forwarding timer expires. Packet duplication is avoided in FSR. We note that packets may be lost during SRNC relocation for these three approaches. Packet loss can not be avoided in SRNC relocation if we want to support real-time applications. We will show that packet loss for FSR is not a serious problem in the following section.

IV. PERFORMANCE EVALUATION

As described in the previous section, the routing path of the downlink packets for the UE is switched from the old path <GGSN→SGSN1→source RNC→target RNC> to the new path <GGSN→SGSN2→target RNC> after the GGSN receives the *Update_PDP_Context_Request* message (Step 5 in Figure 5 (b)). The packets delivered through the old path are lost if these packets arrive at the source RNC later than the *Relocation_Command* message does (Step 7 in Figure 5 (b)). Therefore, an important performance measure is the expected number of lost packets $E[N_L]$ during SRNC relocation. Furthermore, the packets transmitted through the new path are buffered at the target RNC if they arrive at the target RNC earlier than the *Relocation_Commit* message does (Step 8 in Figure 5 (c)). Hence another important performance measure is the expected number of buffered packets $E[N_B]$ during SRNC relocation.

Figure 6 denotes the transmission delays among the net-

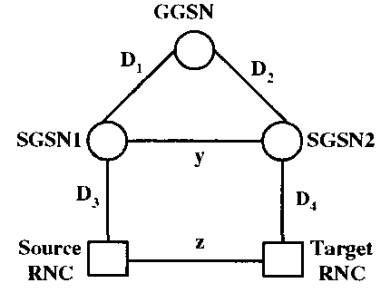


Fig. 6. The Transmission Delays

work nodes, which are represented by the random variables described below.

- D_1 : the transmission delay between the GGSN and SGSN1
- D_2 : the transmission delay between the GGSN and SGSN2. Without loss of generality, we assume that D_1 and D_2 have the same distribution.
- D_3 : the transmission delay between SGSN1 and the source RNC
- D_4 : the transmission delay between SGSN2 and the target RNC. Without loss of generality, we assume that D_3 and D_4 have the same distribution.
- x : $D_2 + D_3$, $D_1 + D_3$ or $D_2 + D_4$.
- y : the transmission delay between SGSN1 and SGSN2
- z : the transmission delay between the source RNC and the target RNC

Based on the above random variables, we develop an analytic model¹ to derive the expected numbers of lost and buffered packets for FSR. Our analytic model has been validated against the simulation experiments. The simulation model follows the approach we developed in [10], and the details are omitted. In our analytic model, we assume that x , y and z have mixed-Erlang density functions

$$f_x(t) = \sum_{j=1}^J \alpha_{x,j} \left[\frac{(\lambda_{x,j} t)^{m_{x,j}-1}}{(m_{x,j}-1)!} \right] \lambda_{x,j} e^{-\lambda_{x,j} t},$$

$$\text{where } \sum_{j=1}^J \alpha_{x,j} = 1$$

$$f_y(t) = \sum_{l=1}^L \alpha_{y,l} \left[\frac{(\lambda_{y,l} t)^{m_{y,l}-1}}{(m_{y,l}-1)!} \right] \lambda_{y,l} e^{-\lambda_{y,l} t},$$

$$\text{where } \sum_{l=1}^L \alpha_{y,l} = 1$$

¹For details of our analytic model, the reader is referred to <http://www.csie.ntu.edu.tw/~acpang/TR.htm>.

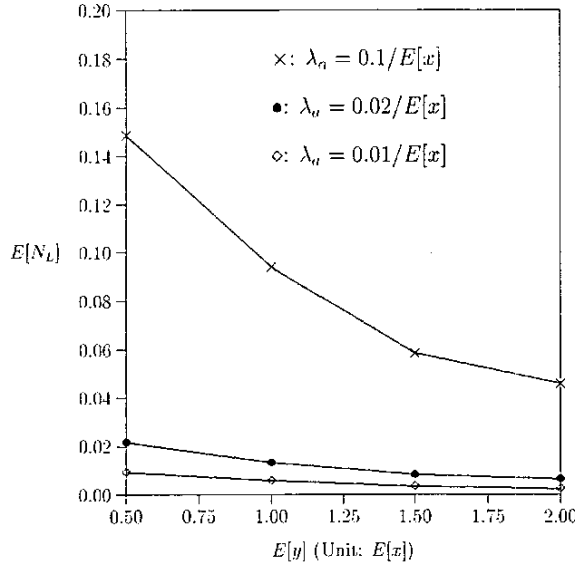


Fig. 7. The $E[N_L]$ Performance (x and y are mixed-Erlang distributed)

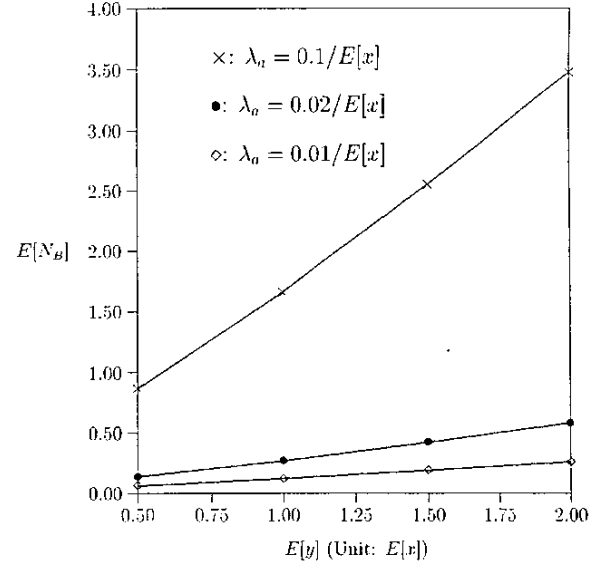


Fig. 8. The $E[N_B]$ Performance (x , y and z are mixed-Erlang distributed)

$$f_z(t) = \sum_{p=1}^P \alpha_{z,p} \left[\frac{(\lambda_{z,p} t)^{m_{z,p}-1}}{(m_{z,p}-1)!} \right] \lambda_{z,p} e^{-\lambda_{z,p} t},$$

where $\sum_{p=1}^P \alpha_{z,p} = 1$

The mixed-Erlang distribution is selected because this distribution has been proven as a good approximation to many other distributions as well as measured data [7], [9]. We also assume that the inter-packet arrivals are a Poisson stream, and the inter-packet arrival times are exponentially distributed with the arrival rate λ_a .

We use numerical examples to investigate the performance of $E[N_L]$ (i.e., the expected number of lost packets) and $E[N_B]$ (i.e., the expected number of buffered packets) for the FSR approach. In our experiments, the mixed-Erlang distributions for x , y , and z have the parameters $\alpha_{x,i} = \alpha_{y,i} = \alpha_{z,i} = 0.5$ and $m_{x,i} = m_{y,i} = m_{z,i} = 2$ for $i = 1$ and 2 . Similar results are observed for other parameter values, which will not be presented here.

Figure 7 plots $E[N_L]$ as a function of $E[y]$ (i.e., the expected number of y) ranging from $0.5E[x]$ to $2E[x]$. The $E[y]$ value is selected depending on whether the SGSNs are located in the same network or different networks. If the two SGSNs are in the same network, the transmission delay $E[y]$ is the same as that between the GGSN and the SGSN. Thus $E[y] \simeq 0.5E[x]$. If these SGSNs are in the different networks, $E[y] \simeq 2E[x]$ may be appropriate. Depending on the applications being investigated, we consider $\lambda_a = 0.1/E[x]$, $0.02/E[x]$ and $0.01/E[x]$. Note that the

100 Mbps Fast Ethernet and 155.52 Mbps STM-1/ATM have been commonly adopted for Gi (between the GGSN and the SGSN) and Iu (between the SGSN and the RNC). For real-time applications such as VoIP and video streaming services, the packet size typically ranges from 200 to 1500 bytes, and the inter-packet arrival time ($1/\lambda_a$) ranges from 10 to 40 ms. Therefore, our study selects the λ_a values in the range $\left[\frac{0.01}{E[x]}, \frac{0.1}{E[x]}\right]$. Figure 7 shows intuitive results that $E[N_L]$ is an increasing function of λ_a , and is a decreasing function of $E[y]$. We also observe that $E[N_L]$ is more sensitive to $E[y]$ for small λ_a than large λ_a . This figure indicates that the $E[N_L]$ performance is reasonably good. For example, when $E[y] = E[x]$, the expected number of lost packets $E[N_L]$ for VoIP application (i.e., $\lambda_a = 0.01/E[x]$) is 0.006. For video streaming services (i.e., $\lambda_a = 0.1/E[x]$), $E[N_L] = 0.09$. Also, when $E[y]$ increases from $E[x]$ to $2E[x]$, $E[N_L]$ is significantly reduced (i.e., 51%, 52% and 58% reductions for $\lambda_a = 0.1/E[x]$, $0.02/E[x]$, $0.01/E[x]$, respectively). In other words, the FSR performance can be improved by increasing the speed of the “ x ” link over the “ y ” link.

Figure 8 shows the effects of $E[y]$ and λ_a on $E[N_B]$, where $\lambda_a = 0.1/E[x]$, $0.02/E[x]$ and $0.01/E[x]$. In this figure, we consider $E[y] = E[z]$ that ranges from $0.5E[x]$ to $2E[x]$. This figure intuitively indicates that $E[N_B]$ increases as $E[y]$ and λ_a increase. Similar to what we observed in Figure 7, the $E[N_B]$ performance is more sensitive to $E[y]$ for small λ_a than large λ_a . Since the expected number of buffered packets at the target RNC is below 3.5 for all cases considered in our study, it is clear that the packet buffering mechanism does not result in long packet delay (due to queuing).

V. CONCLUSIONS

In 3GPP TR 25.936, *SRNC duplication* (SD) and *core network bi-casting* (CNB) were proposed to support real-time multimedia services in the UMTS all-IP network. Both approaches require packet duplication during SRNC relocation, which significantly consume the system resources. This paper proposed a *fast SRNC relocation* (FSR) approach that provides real-time SRNC switching without packet duplication. In FSR, the packet buffering mechanism is implemented to avoid packet loss at the target RNC. We developed an analytic model to investigate the performance of FSR, which was validated against the simulation experiments. We note that packet loss cannot be avoided during SRNC relocation if we want to support real-time multimedia traffic in the UMTS all-IP network. Our performance study indicated that packet loss at the source RNC can be ignored in FSR. Furthermore, the expected number of buffered packets at the target RNC is small, which does not result in long packet delay. FSR can be implemented in the GGSN, SGSN and RNC without introducing new message types to the existing 3GPP specifications. As a final remark, the FSR approach is a US and an ROC pending patents.

ACKNOWLEDGMENT

This work was sponsored in part by the MOE Program for Promoting Academic Excellence of Universities under grant number 89-E-FA04-1-4, Chair Professorship of Providence University, IIS/Academia Sinica, the Lee and MTI Center for Networking Research/NCTU, National Science Council under contract NSC92-2213-E-002-049, CCL/ITRI under contract T1-92021-6, Microsoft and Intel.

REFERENCES

- [1] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; Architectural for an All IP Network. Technical Report 3G TR 23.922 version 1.0.0 (1999-10). 1999.
- [2] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UTRAN Iu Interface RANAP Signaling for Release 1999. Technical Specification 3G TS 25.413 version 3.4.0 (2000-12). 2000.
- [3] 3GPP. 3rd Generation Partnership Project; Technical Specification Group RAN 3; Handovers for real-time services from PS domain. Technical Report 3G TR 25.936 version 4.0.1. 3GPP. 2001.
- [4] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; General Packet Radio Service (GPRS); Service Description; Stage 2. Technical Specification 3G TS 23.060 version 4.1.0 (2001-06). 2001.
- [5] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; Network Architecture. Technical Specification 3G TS 23.002 version 5.3.0 (2001-06). 2001.
- [6] Bos, L., and Leroy, S. Toward an All-IP-based UMTS System Architecture. *IEEE Network*, 15(1):36–45. 2001.
- [7] Fang, Y., and Chlamtac, I. Teletraffic Analysis and Mobility Modeling for PCS Networks. *IEEE Transactions on Communications*, 47(7):1062–1072. July 1999.
- [8] Holma, H., and Toskala, A. (edited). *WCDMA for UMTS*. John Wiley & Sons. 2000.
- [9] Kelly, F.P. *Reversibility And Stochastic Networks*. John Wiley & Sons. 1979.
- [10] Lin, Y.-B., Cheng, H.-Y., Cheng, Y.-H., and Agrawal, P., . Implementing Automatic Location Update for Follow-Me Database using VoIP and Bluetooth Technologies. *IEEE Trans. on Computers*, 51(10):1154–1168. 2002.
- [11] Lin, Y.-B., Huang, Y.-R., Pang, A.-C. and Chlamtac, Imrich. All-IP Approach for Third Generation Mobile Networks. *IEEE Network*, 16(5):8–19. 2002.