3GPP TSG GERAN Tdoc 2g00-072
Adhoc on release 2000 and beyond #1 Agenda item 6.1.2

Helsinki, Finland
August 7th – 11th, 2000

Handover and RAN Relocation for the GERAN

1 Introduction

Several documents have been presented detailing solutions for improvement of hard handover in the UTRAN. The solutions are all derived from the procedures for hard handover in the UTRAN and are of two types:

1. Use of the UTRAN hard handover and SRNS relocation as detailed in 3G.PP 23.060 V3.3 [1]. The primary difficulty with adoption of the existing solution is the imputed difficulty in meeting requirements of handover interruption (150 ms or better for a TCH) during inter-SGSN transfers [2],[3],[4].
2. Improvement of the UTRAN procedures for hard handover through the use of bi-casting. These solutions take two forms:
   a) Bi-casting from the GGSN by establishment of a second GTP tunnel [3],[4].
   b) Bi-casting from the Source RNS [4].

It is generally believed that the current procedures for hard handover and SRNS relocation in 3G.PP23.060 do not meet requirements for real-time traffic. Such a belief has yet to be corroborated by a thorough analysis of the procedure. Indeed, Nokia's document suggests that it is possible to meet requirements for handover interruption even with minimal changes to current baseline [2]. However, it is also a fact that handover interruption times must be comparable to that achievable by GSM in today's networks, and it is allowed that the interruption requirement of 150 ms is easily met by GSM networks in deployment with significant margins[1]. Therefore, Ericsson agree that the current baseline would benefit from enhancements to the procedure detailed in 3G.PP 23.060.

In particular, Ericsson believe that considerable benefits are afforded by the use of bi-casting as a means to reduce interruption time. However, there are specific disadvantages with establishment of a second GTP tunnel from the GGSN. The primary disadvantage of that solution is the fact that the GGSN may be geographically removed from the local market, to the extent that significant delays would occur in the establishment of the second GTP tunnel. Moreover, the establishment of a second tunnel would stress the inter-GSN backbone capacity, especially if a single GGSN site serves the operator's entire network. The other obvious place to establish the bi-casting is the SGSN, but significant changes will have to be made to the Iu interface and to the MM state machine if the SGSN were used to bi-cast user traffic to two GERANs.

The last possible site for bi-casting lies at the BSC in the GERAN, and at the RNC in the UTRAN. As pointed out by Alcatel, this method results in two separate interruptions of downlink user traffic, once when the user hands over, and the other during re-establishment of the GTP tunnel to the GGSN when the target RAN takes over the responsibilities of the Source RAN. Still, the total interruption for the procedure will be within the requirements, with sufficient margin of safety.

[1] Of course, specific figures for handover interruption are proprietary, and it would be difficult for any vendor to provide exact analyses of capabilities in products.

Therefore, this paper proposes that Solution 2(b) above be adopted as an enhancement for hard handover in the 3G network, for the GERAN and the UTRAN. As the solution is one of the options suggested by Alcatel in their paper, this paper is an endorsement of one of the Alcatel approaches. The document also describes the procedure to be followed for relocation of the BSC in the GERAN and the RNC in the UTRAN. The paper details the inter-SGSN relocation procedure. The reduction to simpler cases such as intra-RAN and intra-SGSN handovers is considered obviously derivable. The reduction to a RNC/BSC relocation is also assumed obvious.

2 Assumptions

The following assumptions are made:

- The total handover interruption should be less than 150 ms for all normal cases[2].
- Release 2000 will support handover for real-time traffic served by dedicated physical sub-channels. Shared traffic may be handed over using cell reselection procedures. Support for handover of shared traffic and non-real-time flows is relegated to Release 2001.
- When handover is supported, all RABs corresponding to the MS will be relocated to the new interface.
- Handover commands are sent through the FACCH.
- Measurement reports are sent using the SACCH.
- Support of command and measurement functionality for shared physical bearers is FFS.
- We assume that there is a necessity to perform handovers in the absence of an Iur interface in the GERAN. However, it will be possible to establish a GTP tunnel from one GERAN to another; this capability is used to establish the bi-casting of downlink data traffic.

[2] The interruption requirements may be exceeded by error conditions, and analyses of these are FFS.

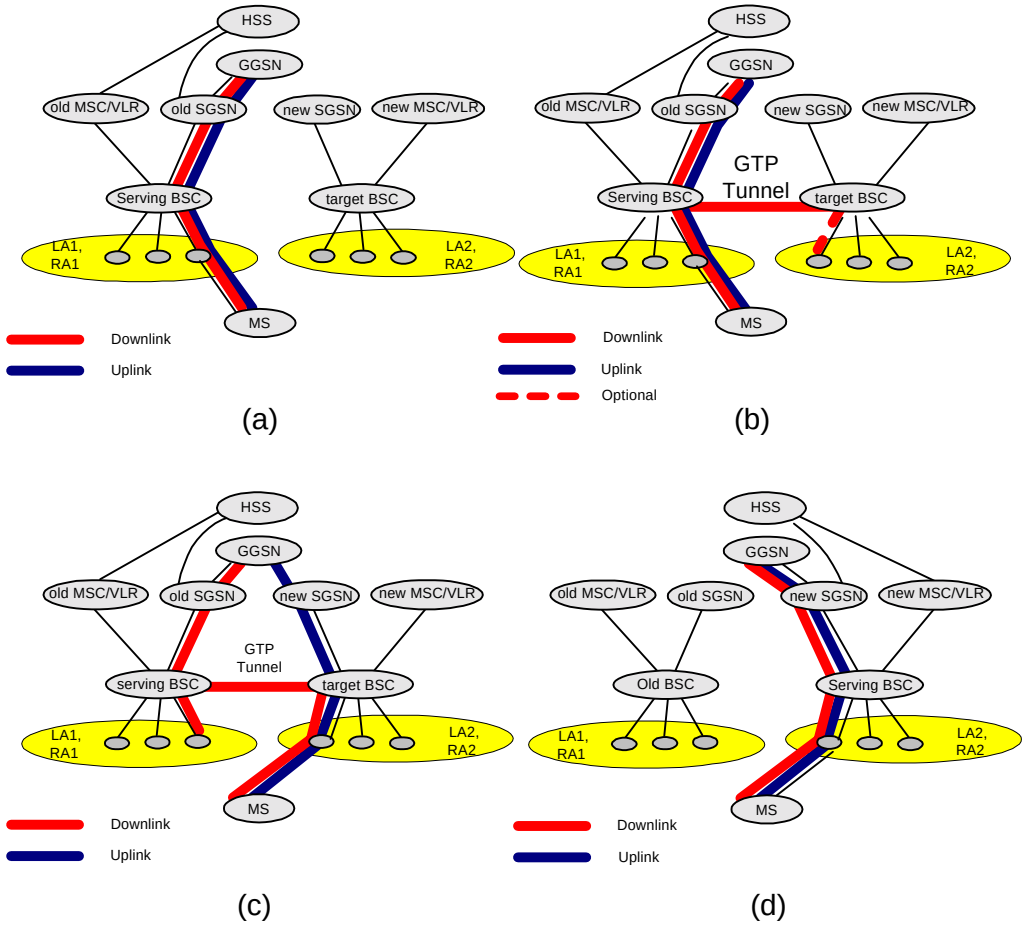

Figure 1. The Architectural flow diagram shows the sequence of operations for an inter-SGSN handover; the operations are described in the text of Section 3.1.

3 Procedure for Inter-SGSN hard handover

3.1 Architectural Flow

The diagrams in Figure 1~~Figure 1~~(a)-(d) detail the procedure to be followed from an architectural perspective. This procedure is detailed in very rough terms below:

a) A handover candidate is identified by the source RAN, and a *Relocation Required* message passed to the SGSN which forwards the *Relocation Request* to the target RAN via the target SGSN.

b) A duplicate GTP tunnel is set up by the source RAN to the target RAN. The tunnel is established in the downlink direction initially and resources are set up for the uplink. A handover command is sent to the mobile.

c) On detection of the handover, the downlink traffic is routed from the GGSN to the target BSC/RNC via the old SGSN and source BSC/RNC. Uplink traffic is routed to the GGSN via the new SGSN, using the existing tunnel identities. The resources in the old cell are released.

d) The GTP tunnel is rerouted from the GGSN via the new SGSN and the target RAN takes up the role of the serving RAN. The HSS is updated with the new LA and RA.

The role of the BSC is taken up by the RNC in the UTRAN. In the treatment that follows, the term RAN is used to indicate GERAN and UTRAN

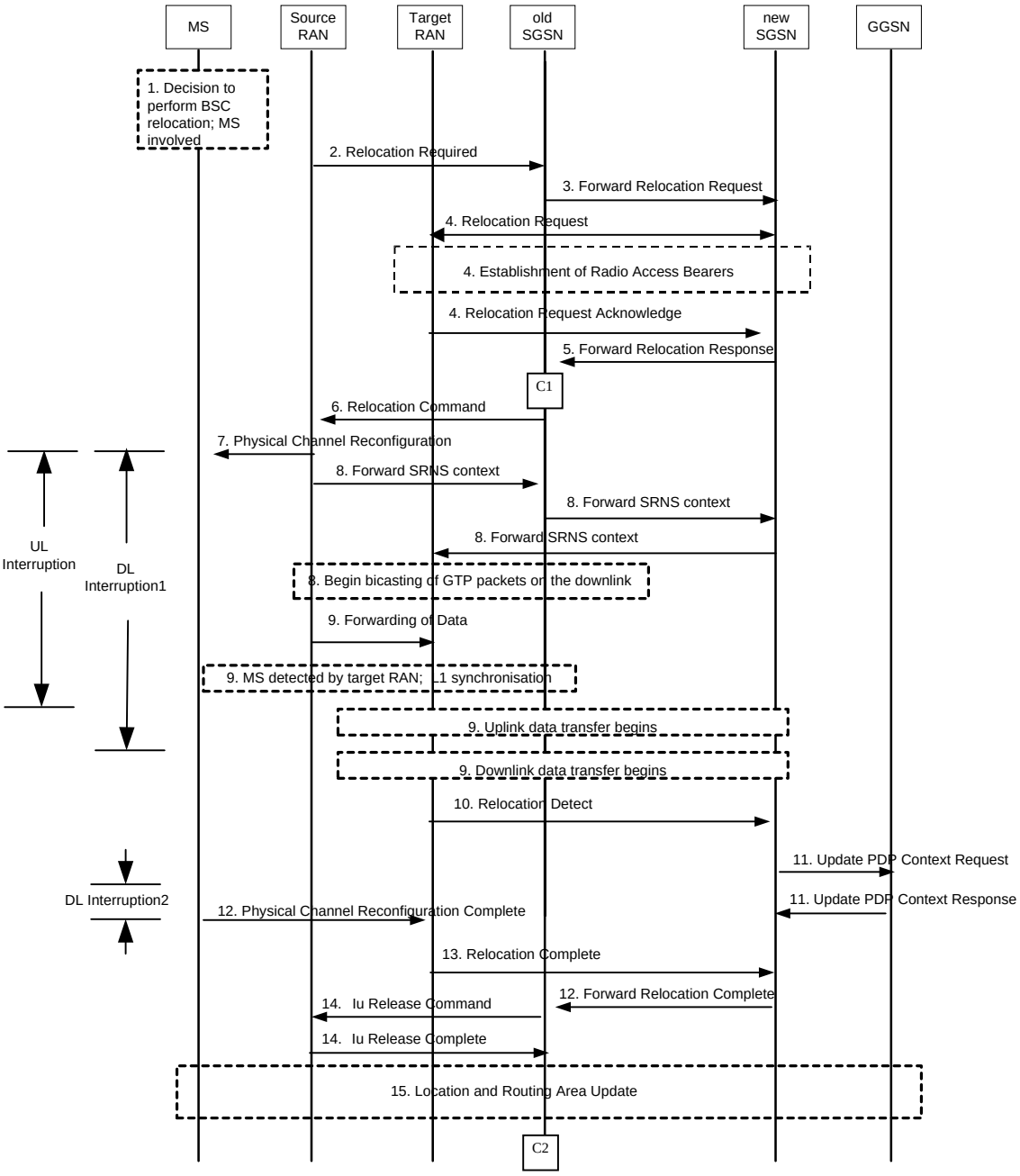

Figure 2. The sequence diagram of 23.060 is modified thus for hard handover combined with BSC/RNC relocation.

3.2 Message Sequence Flow

The message sequence flow of Figure 2~~Figure 2~~ is described below:

1) Based on measurement results and knowledge of the RAN topology, the source RAN (S-BSC/S-RNC) decides to initiate a combined hard handover and RNS relocation.
2) The source RAN initiates the relocation preparation procedure by sending a *Relocation Required (Relocation Type, Cause, Source ID, Target ID, Source RAN To Target RAN Transparent Container)* message to the old SGSN. The source RAN shall *set Relocation Type* to *"MS Involved"*. Source To Target RAN Transparent Container includes the necessary information for relocation co-ordination, security functionality and RRC protocol context information (including MS Capabilities).
3) The old SGSN determines from Target ID if the RAN relocation is intra-SGSN RAN relocation or inter-SGSN RAN relocation. In case of an inter-SGSN RAN relocation the old SGSN initiates the relocation resource allocation procedure by sending a *Forward Relocation Request (IMSI, Tunnel Endpoint Identifier Signalling, MM Context, PDP Context, Target Identification, RAN Transparent Container, RANAP Cause)* message to the new SGSN. At the same time a timer is started on the MM and PDP contexts in the old SGSN (see Routeing Area Update procedure in subclause "Location Management Procedures (UMTS Only)"). The *Forward Relocation Request* message is applicable only in case of inter-SGSN SRNS relocation.
4) The new SGSN sends a *Relocation Request (Permanent NAS MS Identity, Cause, CN Domain Indicator, Source RAN To Target RAN Transparent Container, RABs To Be Setup)* message to the target RAN. For each RAB requested to be established, *RABs To Be Setup* shall contain information such as *RAB ID, RAB parameters, Transport Layer Address, and Iu Transport Association.* The *RAB ID* information element contains the *NSAPI value,* and *the RAB parameters* information element gives the QoS profile. The *Transport Layer Address* is the SGSN Address for user data, and the *Iu Transport Association* corresponds to Tunnel Endpoint Identifier Data.

   After all the necessary resources for accepted RABs including the Iu user plane are successfully allocated, target RAN shall send the *Relocation Request Acknowledge (Target RAN To Source RAN Transparent Container, RABs Setup, RABs Failed To Setup)* message to the new SGSN. The transparent container contains all radio-related information that the MS needs for the handover, i.e., a complete RRC message (e.g., Physical Channel Reconfiguration) to be sent transparently via CN and source RAN to the MS.
5) RANRANWhen resources for the transmission of user data between target RAN and new SGSN has been allocated and the new SGSN is ready for relocation of SRNS, *the Forward Relocation Response (Cause, RAN Transparent Container, RANAP Cause, Target RAN Information)* message is sent from the new SGSN to the old SGSN. **This message indicates that bi-casting of packets from the source RAN should commence**, i.e., the relocation resource allocation procedure is terminated successfully. *RAN transparent container and RANAP Cause* are information from the target RAN to be forwarded to the source RAN. *The Target RAN Information,* one information element for each RAB to be setup, contains *the RAN Tunnel Endpoint Identifier* and *RAN IP address* for data forwarding from source RAN to target RAN. The *Forward Relocation Response* message is applicable only in case of inter-SGSN RAN relocation.
6) The old SGSN continues the relocation of RAN by sending *a Relocation Command (Target RAN To Source RAN Transparent Container, RABs To Be Released, RABs Subject To Data*

Forwarding, RABs Subject to Bicasting) message to the source RAN. The old SGSN decides the RABs to be subject for data forwarding or RABs subject to bicasting based on QoS, and those RABs shall be contained in RABs subject to data forwarding. For each RAB subject to data forwarding or bicasting, the information element shall contain *RAB ID, Transport Layer Address, and Iu Transport Association. Transport Layer Address and Iu Transport Association* is used for forwarding of DL N-PDU from source RAN to target RAN.

7) Upon reception of the *Relocation Command* message from the PS domain, the source RAN shall start the data-forwarding timer. When the relocation preparation procedure is terminated successfully and when the source RAN is ready, then the source RAN shall trigger the execution of relocation of RAN by sending to the MS the RRC message provided in the Target RAN to source RAN transparent container, e.g., *a Physical Channel Reconfiguration (MS Information Elements, CN Information Elements)* message. *MS Information Elements* include among others new RAN identity and S-RNTI. CN Information Elements contain among others Location Area Identification and Routeing Area Identification.
8) The source RAN continues the execution of relocation of the RAN by sending *a Forward SRNS Context (RAB Contexts)* message to the target RAN via the old and the new SGSN. The purpose of this procedure is to transfer SRNS contexts from the source RAN to the target RAN when handover is made with switching in CN. *SRNS contexts* are sent for each concerned RAB and contain the sequence numbers of the GTP PDUs next to be transmitted in the uplink and downlink directions and the next PDCP sequence numbers that would have been used to send and receive data from the MS. For connections using RLC unacknowledged mode PDCP sequence numbers is not used. **Bicasting of data is begun towards the target RAN and the existing PDCP/RLC/MAC context for those RABs that were configured for bicasting.**
9) After having sent the *Forward SRNS Context* message, source S-BSC/S-RNC begins the forwarding of data for the RABs to be subject for data forwarding. The data forwarding at SRNS relocation shall be carried out through the Iu interface, meaning that the data exchanged between source S-BSC/S-RNC and target BSC are duplicated in the source SRNC and routed at IP layer towards the target BSC/RNC.

   The MS is detected by the target RAN. After synchronization is complete, the MS can automatically start receiving downlink transmissions for real-time data**. The MS can start transmission of real-time data to the target RAN. Uplink transmission then proceeds towards the GGSN via the new SGSN, using tunnel end-point coordinates of the GTP tunnel established between the old SGSN and the GGSN.** If needed, additional bearers are used to complete set up of PDCP sequence numbers (PDCP_SNU, PDCP_SND) for RABs designated for data forwarding.
10) The target RAN shall send a *Relocation Detect* message to the new SGSN when the relocation execution trigger is received. For RAN relocation type "UE or MS Involved", the relocation execution trigger may be received from the Um or the Uu interface; i.e., when target RAN detects the UE/MS on the lower layers. When Relocation Detect message is sent, target RAN shall start RAN operation for non-real-time RABs that were handed over with the real-time RABs.
11) Upon reception of Relocation Detect message, CN may switch the GTP tunnel from the GGSN to the new SGSN. If the SRNS relocation is an inter SGSN SRNS relocation, the new SGSN sends Update PDP Context Request (New SGSN Address, SGSN Tunnel Endpoint Identifier, QoS Negotiated) message to the GGSNs concerned. The GGSNs update their PDP context fields and return an Update PDP Context Response (GGSN Tunnel Endpoint Identifier) message. This message is forwarded to the old SGSN and to the new RAN.

12) When the MS has reconfigured it self, it sends e.g., *a Physical Channel Reconfiguration Complete* message to the target S-BSC/S-RNC. From now on the exchange of packets with the MS can start for any remaining RABs configured after handover.
13) The target RAN shall initiate *Relocation Complete* procedure by sending *the Relocation Complete* message to new SGSN. The purpose of *Relocation Complete* procedure is to indicate by the target RAN the completion of relocation of SRNS to the CN. If the user plane has not been switched at Relocation Detect, the CN shall upon reception of Relocation Complete switch the user plane from source RAN to target RAN. If the RAN Relocation is an inter-SGSN RAN relocation, then the new SGSN signals to the old SGSN the completion of the RAN relocation procedure by sending a Forward Relocation Complete message.
14) Upon receiving the *Relocation Complete* message or if it is an inter-SGSN RAN relocation; the old SGSN sends an Iu Release Command message to the source RAN. When the RAN data-forwarding timer has expired the source RAN responds with an Iu Release Complete message.
15) After the MS has finished the reconfiguration procedure and if the new Location/Routeing Area Identification is different from the old one, the MS initiates the Location or Routeing Area Update procedure respectively. See subclause "Location Management Procedures (UMTS Only)". Note that it is only a subset of the RA update procedure that is performed, since the MS is in PMM-CONNECTED state.

For an MS with GPRS-CSI defined, CAMEL interaction may be performed, see referenced procedures in 3G TS 23.078:

C1) CAMEL-GPRS-SGSN-Context-Acknowledge.
C2) CAMEL-GPRS-Routeing-Area-Update.

4 Approximate Analysis of the Voice Handover Interruption

In what follows we provide an approximate analysis of the interruption time for handover of a voice call carried over a dedicated bearer in the GERAN. It is pointed out that the interruption during the period where the mobile is synchronising to the channel is significant in comparison to the time taken for messaging in the network (see the sequence diagram of Figure 2~~Figure 2~~. Accounting for bad frame masking, the interruption due to handover is a maximum of 130 ms on the downlink and 110 ms on the uplink.

The interruption may be increased if the MS has other RABs that have to be configured for relocation. The analysis of the Physical Channel Reconfiguration procedure is for further study. It is expected that the QoS policy function will prioritise bearers for reconfiguration in accordance with the maximum interruption sustainable.

Number	Item	DL?	UL?	DL interrupt (ms)	UL interrupt (ms)
1	Normal voice	Y	Y	0	0
2	Handover command (FACCH)	N	Y	20	0
3	ACK on L2 FACCH	Y	N	0	20
4	Processing(simultaneous with above)	Y	N	0	0
	Bicasting begins (transfer delay)			**0-20**	**0**
5	Mobile resynch	N	N	10-50	10-50
6	Unsynchronized handover (time alignment)	N	N	90-100	90-100
	Uplink transfer loss			**0**	**0-20**
7	Establish layer 2 (1 FACCH frame)	N	N	20	20
8	Establish layer 2 (ACK 1 FACCH frame)	N	Y	20	0
9	HO complete (1 FACCH frame)	Y	N	0	20
10	Normal voice	Y	Y	0	0
11	**PDP context update**	**Y**	**N**	**0-20**	**0**
	Bad frame masking of speech			**80-120**	**80-120**
	Interruption counting possible talk bursts as silence			**200-330**	**200-270**
	Interruption counting possible talk bursts as voice			**160-250**	**160-230**
	Maximum interruption accounting for bad frame masking of 80 ms (only the contiguous interruption period is used for this calculation)			**130**	**110**

Table 1. A rough analysis of the handover interruption through the use of bi-casting from the RAN. It is pointed out that practical implementations will likely achieve figures closer to the higher end of the range than towards the middle of the range shown. The requirement of 150 ms allows a sufficient factor of safety to the above calculations. Network messages take relatively small periods. The Handover command in Point 2 corresponds to Physical Channel Reconfiguration. Bicasting is started after Point 4 in the table and may cause a transfer delay of 1 speech frame that will add up to 20 ms of perceptible interruption to the figures in the table. At Point 9, the uplink flow is active up to the GGSN. The GTP tunnel on the downlink is switched at Point 11, and will cause an imperceptible loss of at most one speeech frame.

5 Conclusions

Ericsson have endorsed the basic concept for real-time handover based on bi-casting from the RAN detailed by Alcatel for the UTRAN in [4]. It is reiterated that the method detailed will yield handover interruption times comparable to GSM systems in deployment. Handover interruptions using the procedure are expected to be less than 150 ms for a single active voice bearer.

Lossless handovers will use a subset of the method in Section 3.2, where RABs are suspended and set up for packet forwarding. The exchange of PDCP sequence numbers will be used for resynchronising packet counters on the new cell. The method is identical to the hard handover procedure detailed in 3G.PP 23.060 today.

Changes are pending in 3G.PP for v.3.4.0 of 23.060. These changes will be incorporated into this contribution when that release is available. It is Ericsson’s intention to submit a version of this paper in 3G.PP as part of the UMTS standards.

References

[1] 3G.PP 23.060, “General Packet Radio Service; Service Description (Stage 2),” v.3.3.0
[2] Nokia, “Considerations for Handover of Real time services from PS domain,” TSG-RAN Working Group 3 meeting #13,” Oahu, Hawaii, May 2000.
[3] Nortel Networks, “Impacts of bi-casting proposal on RAN3 specifications, TSG-RAN Working Group 3 meeting #13,” Oahu, Hawaii, May 2000.
[4] Alcatel, “Relocation for Real Time Services in PS domain,” TSG-RAN Working Group 3 meeting #13,” Oahu, Hawaii, May 2000.