IST-4-027756 WINNER II

D4.8.3

Integration of cooperation on WINNER II System Concept

Contractual Date of Delivery to the CEC: 30/11/2007	
Actual Date of Delivery to the CEC: 29/11/2007	
Author(s):	Emilio Mino[1], Albena Mihovska[2], Bayu Anggorojati[2], Satya Ardhy Wardana[2], Lin Hui[3], Christian Mensing[4], Elias Tragos[5], Jijun Luo[6], Xiaoyun Xue[7], Sana Horrich[7], Roberta Fracchia[8], Annika Klockar[9], Alfonso Tierno[1]
Participant(s):	*TID[1], AAU[2], CATR[3], DLR[4], NTUA[5], SM[6], FT[7], MOT[8], CTH/KAU[9]*
Workpackage:	*WP4*
Estimated person months:	35 PM
Security:	**Public**
Nature:	**Report**
Version:	1.0
Total number of pages:	102

Abstract: This report presents the layered protocol stack of the WINNER wireless system, and describes the system functions associated with the IP Convergence Layer (IPCL), the Radio Resource Control (RRC). The document provides a consolidated description of the system functions associated to the IPCL and RRC layers of the WINNER system concept.
An innovative and hybrid and scalable approach is proposed for the radio resource management architecture It includes a baseline of pure distributed RRM architecture, where many functions reside in the BS (handover, admission control, load control, etc.), and the option of a centralized RRM server, that supports advanced functions that need a central node to coordinate sets of BSs (inter-cell frequency reuse, inter-cell load balancing, etc). The centralized RRM is proposed to obtain optimal performance in situations of medium-high load in the network and also when advanced functionalities are needed.
The combination of WINNER stand-alone location information with the measurecments based on satellite navigation systems (GPS, Galileo) has been analyzed for the location service support. The results have been validated by simulations.
The last part of this document includes the simulation results of the algorithms for intramode, intermode and inter-RAN handovers, using a multimode/multi-RAN terminal. As advanced technique, simulation results on the performance of fuzzy logic based inter-system handover are presented.

Keyword list: IP Convergence Layer, Radio Resource Control, Logical node architecture, System functions, Distributed and Centralised RRM, Radio Paging, Radio and IP Handover, Location determination, QoS based management, Intramode, intermode and intersystem handover simulation, macromobolity, Admission Control, Load Control

Disclaimer:

Executive Summary

WINNER is a B3G wireless network proposal, which is based on evolutionary and revolutionary design approaches, in architectures and system functions. One of the principal motivations of this proposal is the optimization of data rate offered by the network, to facilitate user plane traffic. All the traffic, inside WINNER RAN, will be IP based, disappearing transcoding delays with other protocols (e.g. ATM to IP in UMTS), with a reduced number of nodes to decrease internode user plane signalling, (e.g in UMTS between the node B and the Internet, the traffic has to pass through the RNC, SGSN and GGSN modules, in comparison, in WINNER, the number of entities has been reduced, to have only the GW as the only intermediate node beyond the BS to route the traffic to the core network.)

The WINNER architecture is based on a pure distributed RRM architecture (there is no central RRM coodination by an RNC), where many functionalities have been moved to the BS (handover, admission control, etc.), but some advanced functionalities have been identified that could be lost without a central RRM coordination The use of an optional centralised RRM server has been proposed to offer superior performance in situations of medium-high load in the network, when central coordination performance will provide system functionalities and gains is not achievable by a pure distributed RRM, but withouth disturb user plane traffic. A hybrid and scalable architecture is proposed to the operator.

Radio resource control (RRC) functions at the control plane take the user plane optimization into account. For example, the handover can be triggered by loss of coverage as baseline, like in current systems, but handover can be triggered by better data rate offered by a neighbour cell (see simulations section) has also been introduced. The performance of intermode and intersystem handover algorithms has been validated by simulations.

One highlight of the innovation presented in this deliverable is the two-level Admission Control. Due to the high traffic capability of the radio interface, congestion could be located at radio and now also at backhaul level and these two interfaces should be asked, before admitting new calls or sessions. The proposed solution for intersystem handover is based on the use of fuzzy logic for intersystem handover, allowing the comparison of radio and traffic of heterogeneous networks.
Other innovative functions are the scalable context transfer on handover following the policy-based mobility management. In addition, in order to support load balancing between GWs which targets at a cost effective GW deployment, the hybrid handover mechanism combing IP and radio handover is presented. Furthermore, the location based handover offers the optimal added-value connectivity and service provisioning.

For location determination in previous documents approaches to obtain TDOA timing information by OFDM synchronization algorithms were presented. Here, this WINNER stand-alone location information is combined with satellite system (GPS, Galileo) based positioning techniques. Additionally, for further improvement tracking algorithms for the solution of the navigation equation in the dynamic case were applied which helps to improve the estimates in average. Furthermore, emergency calls, inter-system handover, and radio resource management were identified and analyzed as system-side applications that can exploit available location information.

Table of Contents

Authors

Partner	Name	Phone / Fax / e-mail
AAU Aalborg University	Albena Mihovska	Phone: +45 96358639 Fax: +45 98151583 e-mail: albena@kom.auc.dk
AAU Aalborg University	Bayu Anggorojati	Phone: +45 96358647 Fax: +45 98151583 e-mail:ba@es.aau.dk
AAU Aalborg University	Satya Ardhy Wardana	Phone: +45 96358647 Fax: +45 98151583 e-mail: saw@es.aau.dk
CATR China Academy of Telecommunication Research of MII	Lin Hui	Phone: +86 10 62304576 Fax: +86 10 62304471 e-mail: linhui@mail.ritt.com.cn
CTH/KAU Chalmers tekniska högskola/Karlstads universitet	Annika Klockar	Phone: +46 0709754670 Fax: +46 547001828 e-mail: Annika.Klockar@kau.se
DLR Deutsches Zentrum für Luft- und Raumfahrt	Christian Mensing	Phone: +49 8153 282878 Fax: +49 8153 282875 e-mail: christian.mensing@dlr.de
FT France Telecom R&D	Xiaoyun Xue	Phone: +33 1 45294564 Fax: +33 1 45294194 e-mail: xiaoyun.xue@orange-ftgroup.com
FT France Telecom R&D	Sana Horrich	Phone: +33 1 45296583 Fax: +33 1 45294194 e-mail: sana.horrich@orange-ftgroup.com
MOT Motorola Lab	Roberta Fracchia	Phone: +33 1 693548466 Fax: +33 1 69354801 e-mail: roberta.fracchia@motorola.com
NTUA National University of Athens	Elias Tragos	Phone: +302107721511 Fax: +302107722534 e-mail: etragos@telecom.ntua.gr

NSN Nokia Siemens Networks	Jijun Luo	Phone: +49 89 636 45268 Fax: + 49 89 636 45591 e-mail: jijun.luo@nsn.com

TID Telefónica I+D	Alfonso Tierno	Phone: +34 91 3374863 Fax: + +34 91 3374212 e-mail: atierno@tid.es

TID Telefónica I+D	Emilio Mino	Phone: +34 91 3374428 Fax: + +34 91 3374212 e-mail: emino@tid.es

List of Acronyms and Abbreviations

3GPP	3G Partnership Program
AC	Admission Control
ACK	Acknowledge
AES	Encriptation algorythm
AICH	Acquisition Indicator Channel
AOA	Angle of Arrival
AP	Access Point
ARQ	Automatic Repeat Request
AS	Access Stratum
AuC	Authentication Center
B3G	Beyond 3G
BER	Bit Error Ratio
BCCH	Broadcast Channel
BLER	Block Error Rate
BS	Base Station
BSla	Base Station local area
BSma	Base Station metropolitan area
BSwa	Base Station wide area
CAPEX	Capital Expenditure
CDF	Cumulative Density Function
C/I	Carrier to Interference
CP	Control Plane
DL	Down-Link
DRX	Discontinous Reception
E2E	End to End
EKF	Extended Kalman Filter
FCC	Frame Control Channel
FDD	Frequency Division Duplex
FTP	File Transfer Protocol
GGSN	Gateway GPRS Support Node
GoS	Grade of Service
GNSS	Global Navigation Satellite System
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HARQ	Hybrid ARQ
HCS	Hierarchy Cellular Structure
HIS	Hybrid Information System
HLR	Home Location Register
HO	Handover
HRRM	Hybrid RRM
HSS	Home Subscriber Server
ID	Identification
IMS	IP Multimedia System
IP	Internet Protocol
IPCL	IP Convergence LAyer
ISHO	Inter-System Handover
Iu-Flex	3GPP interface
KF	Kalman Filter
LA	Local Area (deployment concept)
LDCCH	Logical Dedicated Control Channel
LDTCH	Logical Dedicated Transport Channel
LD	Location Determination
LOS	Line of Sight
LTE	Long Term Evolution
MA	Metropolitan Area (deployment concept)

MAC	Medium Access Control
MIP	Mobile IP
MADM	Multiple Attribute Decision Making
MMSE	Minimum Mean Square Error
NAS	Non Access Stratum
NLOS	Non-Line of Sight
NTI	Network Temporary Identity
OFDM	Orthogonal Frequency Division Multiplexing
O&M	Operation and Maintenance
OMC	Operations & Maintenance Centre
PBCH	Physical Broadcast Channel
PDCP	Packet Data Convergence Protocol
PDU	Protocol Data Unit
PER	Packet Error Rate
PHY	PHYsical layer
PPCH	Physical Paging Control Channel
PRACH	Physical Random Access Channel
PLMN	Public Land Mobile Network
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RLC	Radio Link Control
RMSE	Root Mean Square Error
RNC	Radio Network Controller
RRC	Radio Resource Control
RRM	Radio Resource Management
RS	Resource Scheduller
RSS	Received Signal Strength
RTT	Round Trip Time
RTTM	Real Time Measurement
S1-Flex	3GPP interface
SAE	System Architecture Evolution
SAP	Service Access Point
SDU	Service Data Unit
SGSN	Serving GPRS Support Node
SINR	Signal to Interference plus Noise Ratio
SLC	Service Level Controller
SNR	Signal to Noise Ratio
SN	Sequence Number
SRMML	Specific RRM entity in the Legacy RANs
SRRMW	Specific RRM entity in WINNER
TA	Tracking Area
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TDOA	Time Difference of Arrival
TOA	Time of Arrival
TX	Transmission
UDP	User Datagram Protocol
UIA2	Encriptation algorythm
UL	Up-Link
UMTS	Universal Mobile Telecommunication System
UP	User Plane
UT	User Terminal
UTRAN	UMTS Terrestrial Radio Access Network
VoIP	Voice over IP
WA	Wide Area (deployment concept)
WWI	Wireless World Innitiative
WLAN	Wireless Local Area Network

1. Introduction

The project WINNER has defined the principal parts of a complete radio access network of next generation. The studies carried out have supposed the development of innovative solutions in respect to current systems, but keeping compatibility with current systems (namely UMTS and 3GPP LTE). In some cases it was not needed to design a new approach to a system function, it was adopted the solution provided by the legacy wireless systems, but adaptep to the specific WINNER system architecture and protocol layers. Some of the higher layers functions and new solutions provided by WINNER are included in this document (see conclusion section), in particular, those related with the IPCL (IP Convergence Layer) and the RRC layer (Radio Resource Control).
The structure of this document is as follows:

Chapter 1 presents the overall WINNER layered protocol stack and the mapping of the protocol stack in the essential logical nodes; afterwards the functions and signalling of each function are described. The objective of this document is to present the higher layers system functions, which are described considering a layered approach and the involved protocols and signalling between the logical nodes. The document is structured in five parts organised in different technical chapters.

Chapter 2 describes the overall protocol stack and the mapping of the protocol stack in the different basic logical nodes (UT, BS, GW).

Chapter 3 presents the services, functions and protocols of the IPCL protocol layer, as header compression and decompression, transfer of user data between IPCL peer entities (typically the UT and the GW), the in sequence delivery of upper layer PDUs, the duplicate detection of lower layer SDUs, inter GW handover and ciphering of IPCL user plane data (and NAS signaling) .

Chapter 4 describes the system functions and protocols associated to the RRC layer. These have been divided in two groups: system concept functions, and advanced functions.

The system concept functions (the majority of them belong to the RRC layer), are those that the WINNER system needs in order to operate. These functions are listed as:

- Idle UT mobility management. The paging protocol supports UT mobility in idle mode; this section includes the study of the needed changes in the MAC superframe to support this function. Also cell selection and reselection is presented.
- Broadcast of system information. Identification of the information to be broadcasted, at cell level, to UTs, in idle and active modes.
- Admission Control. A new approach for B3G networks is needed, the rational is that due to the high traffic capability of the radio interface, congestion could be located at radio or at the backhaul, and therefore congestion on backhaul should be also considered.
- Establishment, maintenance and release of an RRC connection. The first steps of network access, after the physical layer initial synchronization process are described. Also the interactions, with other nodes outside of WINNER RAN (for purposed as Authentication, Ciphering and Temporal Identites) are drafted.
- Active mode micro-mobility. Mobility management in active state, intramode and inter-mode handover protocols, UT measurement reporting, UT cell selection and reselection, UT context transfer, and RLC SDU context transfer are described.
- Intersystem handover. Final solution considering the WWI functional architecture for the intersystem handover.
- Flow admission. Is the mechanism that receives the requests for new flows (whether they come from a new user or from ongoing users) and checks if the users are authenticated to the network and if the network has sufficient resources based on the requested resources by the new session.
- Load/Congestion Control. Description of flow control phases and its associated interactions (load monitoring, congestion resolution and congestion recovery phase)
- Integrity protection and ciphering of RRC messages. For preventing the insertion and modification or RRC messages and also to keep the confidentiality of the contents.

The advanced functions (the majority of them belong to the RRC layer) that enhance system performance or provide new functions are the following:

- Hybrid and scalable network. The baseline is a distributed RRM architecture, but it is offered a centralized RRM architecture with advanced system functions (described in detail in chapter 5)

- Distributed and centralized handover and hybrid handover (mix of the two types of handover)
- Radio and IP handover. Coordination of these two types of handover.
- Load control. Balance of load for the pool of gateways concept.
- Policy based management.
- RLC PDU context transfer during handover. Context transfer is done in order to make a seamless handover process, i.e. without disturbing the ongoing session and also to make the overall process to be faster, i.e. no need to re-establish and re-authenticate the RLC connection.

The section about location service support analyzes the application of tracking algorithm for the combination of WINNER stand-alone location information with the measurements based on global navigation satellite systems (GPS, Galileo) where for the simulations WINNER channel models for wide area and microcellular scenarios were used.

Chapter 5 is devoted to the WINNER architecture, that has continuously been updated and developed since the earlier referenced versions [WIND4.8.1] [WIND4.8.2], in order to increase the adaptability to different deployment scenarios (Wide, Metropolitan and Local Area) and also considering network performance and cost efficiency. The number of logical nodes has been reduced, to reduce the signalling overhead and increase user plane throughput. The challenge has been how to maintain all the control functionalities available in the previous RRM partially centralised structured architecture, and at the same time to offer high efficiency in data transmission.

Chapter 6 presents the final simulation results for different types of handover: intramode, intermode and intersystem are presented, using algorithms based on: signal strength, cell data throughput and UT velocity, using a multimode/multi-RAT terminal and a common RRM entity that coordinates the different networks.

2. WINNER protcol stack and logical nodes

The overall radio access protocol layers are shown in Figure 2-1, comprising Physical layer (or layer 1), Data Link layer (or layer 2), with the sub-layers MAC, RLC and IPCL and Network layer (or layer 3) that only includes the RRC layer.
The protocol architecture is subdivided in user plane (UP), composed of protocols devoted to data transfer services of the user and control plane (CP) composed of the protocols created to control data transfer, user and network operation. The IPCL and RRC sub-layers only have user plane and control plane respectively. As we can see a user plane connection (an IPCL layer session) can generate several RLC layer flows using different QoS classes.

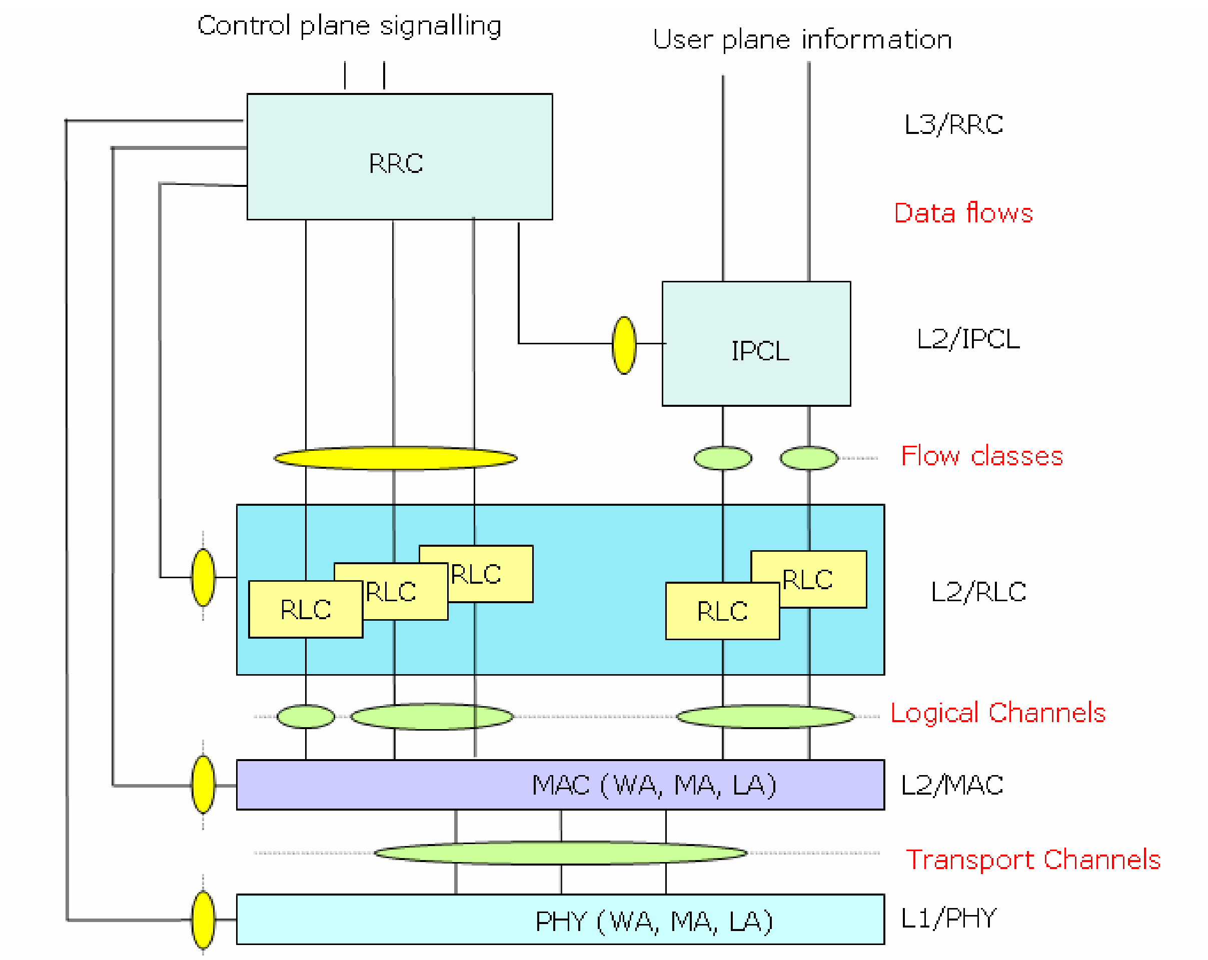

Figure 2-1: Radio Access protocol layers

The communication between consecutive layers and sub-layers is carried out through serving access points (SAPs), green ellipses in Figure 2-1. The communication between the RRC control layer with the other layers is carried through control SAPs (C-SAPs), yellow ellipses in Figure 2-1. Using SAPs and the lower layers offers services to the upper layers.

Over the radio access protocols there is a Non Access Stratum protocol. The NAS control protocol is not fully covered by the scope of this document and is only mentioned for reference of functions that need interact with this protocol (e.g. paging). This document focus principally in IPCL and RRC sublayers.
WINNER is considering two different types of PHY layers, FDD and TDD and three deployment modes associated with three MAC layers, Wide Area (WA), Metropolitan Area (MA) and Local Area (LA)

Figure 2-2 presents the Control Plane protocol layers in the basic elements of the radio access network: UT, BS and GW (RRM server, relays and spectrum server are not included). RRC protocols and services are terminated at the network side at the BS and the Non Access Stratum (NAS) protocols ends at the GW in the WINNER RAN, but continue in the Core Network (HSS) in security and ciphering protocols.

Chapter 4 enumerates and describes the RRC functions. Regarding NAS, it will be involved in idle mode mobility management and paging (in idle mode), authentication, authorization and accounting and also on data flow establishment management and release.

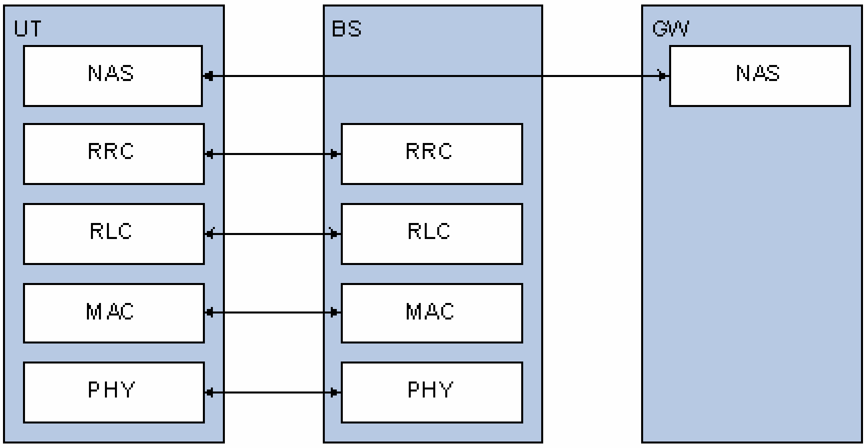

Figure 2-2: Allocation of the radio Control Plane protocols

It is for further study the need of IPCL protocol, in the control plane, for ciphering integrity protection of NAS signalling

Figure 2-3 presents the User Plane protocol layers in the basic elements of the radio access network. The radio access protocols terminated at the network side at the BS, except the IPCL protocol that ends at the GW. IPCL functions and services are described in Chapter 3. In the User Plane, the GW performs the conversion between IPCL (protocol used in the WINNER RAN) and IP/UDP/x (protocols used outside the WINNER RAN). Therefore, over the IPCL layer of the UT and the GW there will be an IP layer that uses IPCL transfer of user data service.

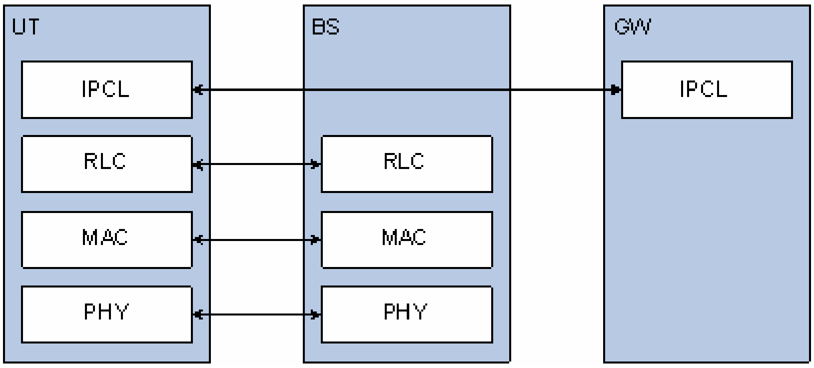

Figure 2-3: Allocation of the radio User Plane protocols

3. IPCL Layer

3.1 IPCL services and functions

IPCL adapts the data flow to the transmission modes of the RLC. Figure 3-1 shows the header added by the IPCL layer to the data flows to create a RLC SDU, composed by an IPCL header and the IPCL SDU. The IPCL layer uses the RLC layer services. These services are the following: transparent data transfer, unacknowledged data transfer and acknowledge data transfer

In the user plane, the minimum data quantity that one layer offer data transference services to its upper layer is called Service Data Unit (SDU) and the minimum date quantity that one upper layer transfer data to a lower layer is called Protocol Data Unit (PDU)

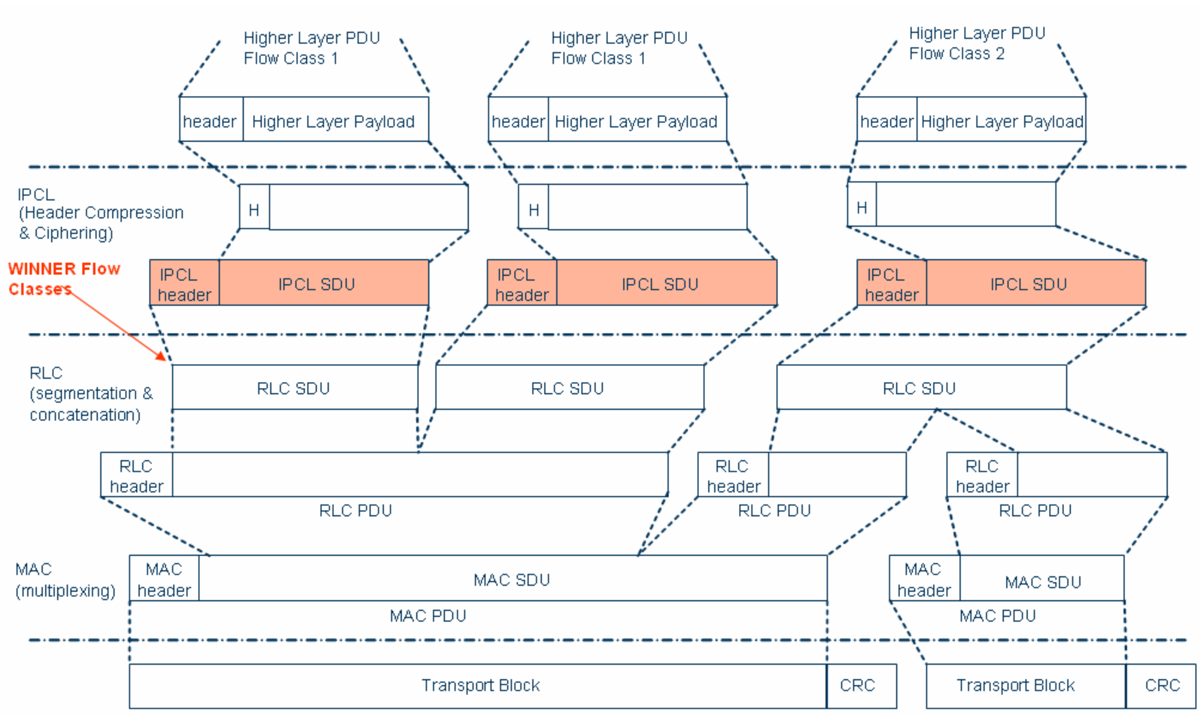

Figure 3-1: RLC SDU and IPCL SDU

The list of services and functions provided by the IPCL layer to the upper layer is the following:

- Header compression and decompression
- Ciphering of user plane data and control plane data
- Integrity protection of control data (U-plane data integrity is covered by the RLC layer)
- Transfer of user data: transmission of user data means that IPCL receives IPCL SDU and forwards it to the RLC layer and vice versa.
- Reordering of the RLC SDUs at least during handover in the DL
  - Performed by the UT, early RLC SDUs not acknowledged and should be retransmitted again by GW
- In-sequence delivery of upper layer PDUs at handover in the UL
  - Performed by the GW, high RLC transmission window, early RLC PDU has to be retransmitted in new BS without acknowledgement
- Duplicate detection of lower layer SDUs
  - Both DL and UL, in case of failure of the ACK upon handover

3.2 Header compression and decompression

The aim of the header compression and decompression protocols of IP data streams (e.g., TCP/IP and RTP/UDP/IP headers) is to reduce the amount of redundant headers information transmitted. IP header has 40 bytes, which can be reduced to only 4 or 7 octets. The first time it is transmitted the full header, and after are transmitted only the changes of this header, but from time to time is send a refresh packet with the full header.

The network layer protocol type, e.g. IP or PPP, is indicated during the flow establishment

Figure 3-2 shows the field of the IPCL Protocol Data Units (PDUs) composed of:

- PDU type. Indicator of IPCL PDU with/without sequence number. 3 bits
- PID (packet identifier). Indicator of the use compression method (or not). The header compression method is specific to the particular network layer (IP, PPP, etc), transport layer (TCP, RTP, UDP, etc.) or upper layer protocol combinations e.g. TCP/IP or RTP/UDP/IP. 5 bits
- Sequence number. Used in the case of flows of compressed packets, with number of sequence (AM mode) .16 bits
- Data. Three types of data
  - Uncompressed IPCL SDU
  - Header compressed IPCL SDU
  - Header compression protocol, including feedback information

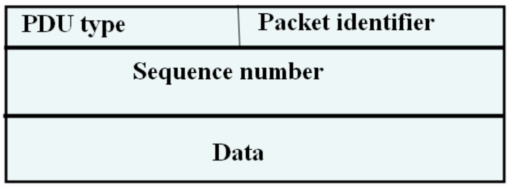

Figure 3-2: IPCL PDU

3.3 Transfer of user data between IPCL entities

Transfer of user data between peer PDCP entities (transmitter and receiver) this service is used by the UT and the GW to transfer IP/UDP packets between them, over this protocol there is an IP/UDP layer in the UT and GW. In Figure 3-3 is shown the transmission protocol, using the acknowledge mode. If the header compression is configured, the sender IPCL entity shall:

- Perform header compression upon reception of IPCL SDU from upper layers
- The Sequence number is incremented in one.
- Submit the IPCL PDU to the RLC layer in sequence received from upper layers

When the receiver IPCL entity receives an IPCL PDU, with compressed header, from the RLC layer, it shall:

- Perform header decompression of the IPCL PDU to obtain IPCL SDU
- To check that is received the expected consecutive sequence number
- Deliver the IPCL PDU to the upper layer, in the order received from lower layer

Two data transfer methods using acknowledged and unacknowledged mode. A typical example of IPCL transmission is an application (the IPCL User in Figure 3-3) downloading data from a server located at Internet.

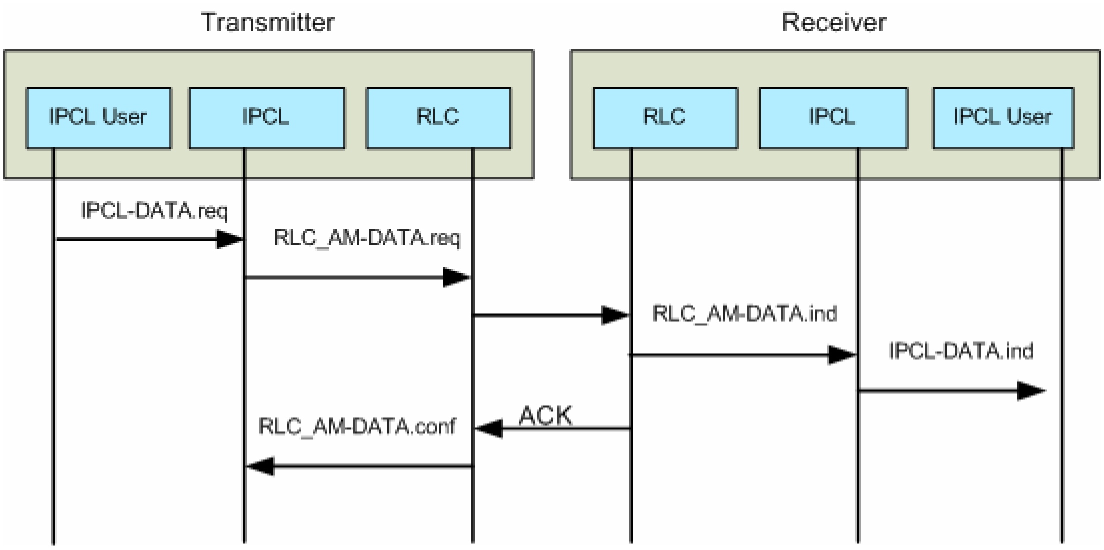

Figure 3-3: IPCL transmission using the RLC services in acknowledge mode

3.4 IPCL and handover

IPCL layer supports non loss of data transfer and efficient data forwarding to higher layers during handover. The efficiency is mainly provided by detection of duplication and in sequency delivery of the PDUs.

In the micro- mobility case, the IPCL performs the following functions:
- Noticing the UL_Receive IPCL SN of the next IPCL SDU expected to be received from the UT. The process is an implicit function which has to be set with relative higher priority during handover. The process can be considered as an alarming preparation while the phase with rather high failure rate compared to the non-Handover phase.
- Same as UL, in the DL, IPCL performs also noticing (alarming preparation) the DL_Send IPCL SN of the first transmitted but not yet acknowledged IPCL SDU.
- Continue to transmit not yet acknowledged SDUs together with their related DL_Send IPCL SNs (the SN is anyway encapsulated in the RLC SDU, the RLC layer does not need to decode the SN)
- Continue to transmit the not yet transmitted IPCL SDUs.
- Duplication detection

More detailed IPCL functions are explained in later sections. The macro mobility case for inter-operator, inter-system and inter GW belonging to the same service pool are very similiar with only one difference reflected by the reset of the SN.

3.5 Inter GW handover

In the macro mobility case, when the anchor point of the U-plane connection are changed, the IP mobility including classic MIP (mobile IP), HMIP (Hierarchy MIP) or FMIP (Fast MIP) can be implemented.

When requested by the RRC layer MM module, for each flow configured to support lossless inter GW Relocation, the IPCL sublayer in the source GW should forward the following to the target GW:
- the UL_Receive IPCL SN of the next IPCL SDU expected to be received from the UT;
- the DL_Send IPCL SN of the first transmitted but not yet acknowledged IPCL SDU
- the transmitted but not yet acknowledged IPCL SDUs together with their related DL_Send IPCL SNs;
- the not yet transmitted IPCL SDUs

IP packets transferred from the corresponding node are not processed by IPCL in the target GW as long as the old packets forwarded by the source GW are not processed. It is recommended to restart the IPCL sequence number at least after the tunnelling phase between the source GW and target one when applicable.

When the inter-domain handover taken place excluding the handover between GWs belonging to the same service area, the inter GW handover does not need to transfer the IPCL context. The losslessness however can be guaranteed by the flow control between GW and BS and issuing the handover command at the right time.

In short, when the inter-GW transport is not secure enough, it is recommendated to discard IPCL context transfer by avoiding vulnerable network caused by security key transfer.

3.6 In sequence delivery of upper layer PDUs

An important function for IPCL is to provide in-sequence delivery of upper layer PDUs. In DL, data may arrive out of order to the target BS, if the data transmitted directly from the GW to the target BS arrive before the data that are forwarded from the source to the target BS during handover (further described in sections 4.1.5.5 and 4.2.6). In UL, the GW may receive data out-of-order, if there are gaps in the data transmitted from the source BS before handover and retransmissions of the missing data arrive from the target BS after handover [Bajzik07].

In 3GPP LTE, out-of-order data may degrade TCP performance severely [Bajzik07, Racz07], because data are delivered out-of-order to TCP. In contrast to IPCL in WINNER, PDCP in 3GPP LTE cannot provide reordering and in-sequence delivery, since it is terminated in the BS and UT.[1] In WINNER, degradation of TCP

[1] In the latest version of [3GPP300], a soultion is proposed. PDCP in the source BS informs the target BS about the next PDCP sequence number to use.

performance due to out-of-order segments should be less of a problem, since IPCL terminates in the GW and UT.

Even if IPCL can reorder data to provide in-sequence delivery to higher protocol layers, out-of-order data should be avoided when possible, since out-of-order data increases delay on higher protocol layers. In [Bajzik07], a separate service class for forwarded data is proposed. To reduce delay of forwarded data, the scheduler in the target BS is proposed to give priority to forwarded data. In [Racz07], forwarded data is proposed to be prioritized over data from the GW in the target BS for transmission to the UT (which will only work if the buffer of forwarded data is not emptied before all forwarded data have been received). We recommend that forwarded data are prioritized in WINNER, in order to avoid out-of-order data.

3.7 Duplicate detection of lower layer SDUs

Data may be duplicated after handover, if a lower layer PDU, MAC or RLC (in case of no PDU context transfer), that is received before handover is delivered to higher layers and the acknowledgement is lost [Bajzik07]. Duplicates may have a negative impact on higher layer protocols, especially on TCP [Bajzik07, Racz07]. In UL, IPCL in the GW should remove duplicates before data is delivered to higher layers.

In DL, there are two alternatives

1. RLC in the target BS uses the IPCL sequence number to detect duplicated data (IPCL PDUs).
2. IPCL in the UT detects and removes duplicated data (IPCL PDUs).

On the one hand, if alternative 1 is used and RLC in the BS performs duplicate detection, then RLC has to look into the IPCL sequence numbers, which violates the protocol layering. On the other hand, if alternative 2 is used and IPCL in the UT performs duplicate detection, then layering is preserved, but duplicates are transmitted all the way over the air to the UT. If there is enough capacity between the BS and the UT, then the UT could detect and remove duplicates and violation of protocol layering could be avoided. Duplicates are not expected to occur often. Therefore, we recommend that IPCL in the UT performs duplicate detection. In environments in which duplicates occur frequently, RLC in the BS could perform duplicate detection as a value added function.

3.8 Ciphering of IPCL data

In the IPCL layer, the ciphering can guarantee the confidentiality for UP traffic. In other words, it ensures that unauthorized users have no access to data belonging to other users. Thanks to ciphering, the eavesdropping of messages, unlawful interception of packets, can be avoided. Moreover, ciphering is also able to ensure to some degree the integrity and authentication as it proves the possession of the corresponding keys[2] shared by the communicating entities. Figure 3-4 depicts the process followed to obtain the IPCL key.

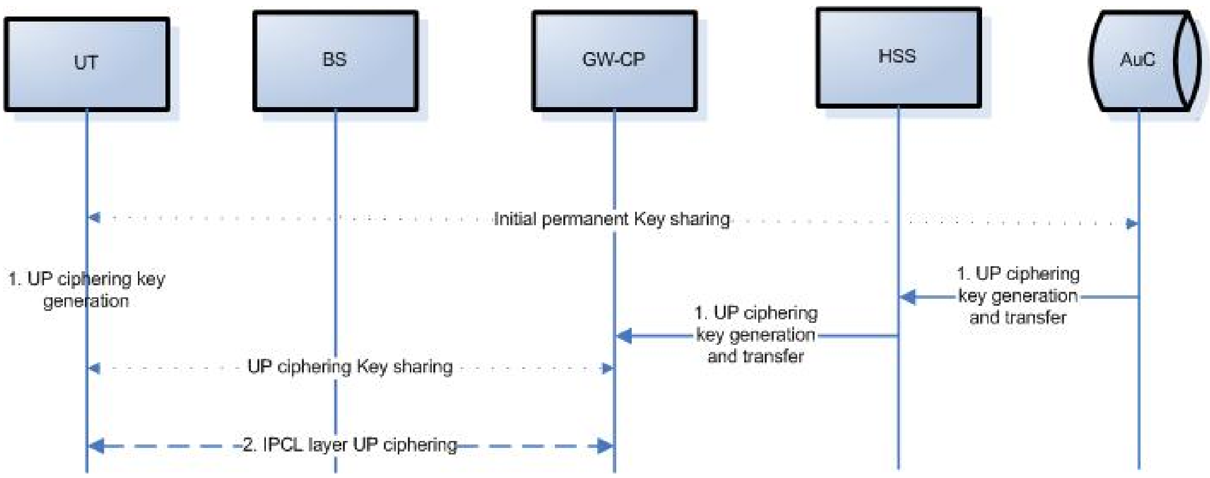

Figure 3-4: Location of security keys in WINNER IPCL layer

[2] Initially, it is supposed that there is a permanent key shared by UT (in the SIM of UT which is supposed to be tamper-resistant) and CN. Alternativelly to have a common key, can be a private/public key, where the private key will be on the UT SIM and the corresponding public key will be only in the HSS data base. After each authentication, a set of ciphering and integrity protection keys will be generated using this permanent key.

The IPCL key will be derived from an intial permanent key stored in the UT SIM and also in the AuC (Authentication Center), which is associated to HSS (Home Subscriber Server). The HSS will choose a random IPCL key, which will be encrypted using the initial common key. The GW will receive the IPCL with and without encriptation; the encrypted key will be transferred by the air to the UT, which will be able to decript the created HSS key. The IPCL will be used by UP traffic and also by the Control Plane NAS messages. Also there will be a second encriptation for the RRC traffic between the BS and UT (see section 4.1.11)

The ciphering algorithms from IPCL and NAS communicactions could be UEA2 and AES. Referring to sections 3.4 and 3.5, the ciphering key are able to be exchanged between GWs in case the connection between GWs is secure enough. That is assumed the inter GW key exchange is only taken place in case the GWs are belonging to the same service pool belonging to the same operator.

4. RRC functions and RRM

In this chapter the most importand radio resource management functions and the related RRC signalling are described. These functions are essential for the operation of a wireless network. The functions that are described include cell selection and reselection, admission control, mobility management, load control, paging, flow control, measurement reporting etc. Figure 4-1 depicts the RRC (and NAS) states and the system information needed in each state.

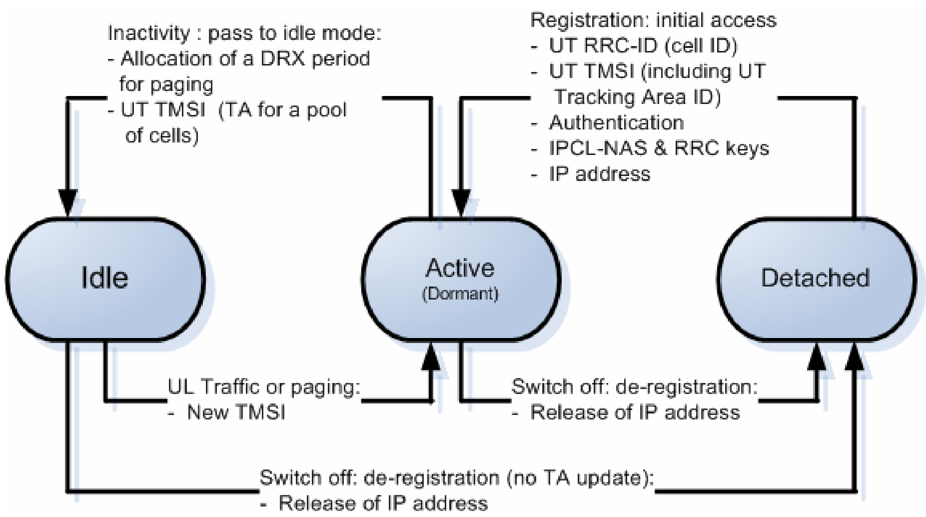

Figure 4-1: RRC states

Connected and detached modes are two basic RRC modes for the UTs. Inside the Connected main state, there will be a separation of the states according to the UT activity, namely 'Active' and 'Idle'. Furthermore, we can distinguish inside the Active mode a submode called 'Dormant'. Therefore, the transition between 'Active' and 'Dormant' is a rather important issue.

The future systems are in favor to have simplified RRC states compared to the UTRAN of UMTS. In order for WINNER to be harmonised and compatible with the 3GPP LTE states, similar RRC states are proposed[3].

These are:

- UT Detached:
    - The network is not aware of this UT (e.g. UT is switched off or not operating on the current system).

- UT Idle:
    - After switched on, the UT camps in the most appropriate cell (for ex. a cell with the highest signal strength). The UT is able to send and receive system information and cell information.

[3] The Winner system differentiates itself to the 3GPP LTE system as: more options in deploying MIMO, Multi-hop using Relay nodes, positioning, Hybrid Information system, Multi-mode UTs which will be supported by more systems simultaneously, etc.

- The UT is maintained in this state until it is going to send or receive user information.
- Inherently power saving state: UT should be able to maintain this state several days.
- UT performs periodic search for higher priority mode or RAN.
- UT performs cell selection and reselection autonomously.
- UT monitors broadcast and paging channels.
- UT is handled by the GW node.
- Tracking area/paging group: UT belongs to a group of cells controlled by a GW (after the initial connection the UT can change its position without notifying the network).

- UT Active (could be similar to the UTRAN CELL_DCH state [3GPP-TS25.331])
  - UT is moved to this state from idle state when transmitting/receiving data on traffic channels.
  - UT monitors the directed/common control channels continuously.
  - UT is handled by BS.
  - Mobility: The UTs select the best cell from the list of candidates provided by network (T8 assumption on mobility)
  - There is a power saving sub-state within the Active state. This is the dormant sub-state where the UT is ready to transmit or receive data on traffic channels and monitors the control channels discontinuously.
- The change from the Active to the Dormant sub-state will be activated when no user plane transmission or reception is expected by the UT. The aim is to reduce power consumption in the active state, with reduced activity cycle. The user plane connection between BS and UT is maintained.
- The change from the Active to the Idle state will be launched by an Active State Timer in the BS, the BS will inform the GW-UP which will then remove the user plane connection between BS and UT.

4.1 System concept functions

4.1.1 Idle mode Mobility Management

In the RRC Idle state (RRC_IDLE), a UT is configured for Discontinuous Reception (DRX) to allow power saving. The UT is allowed to become active once per DRX cycle. In this period, the UT has to become active, and check the PICH to know whether it is currently being paged by the network or not.

To ease the mobility control, fast handover and for security reasons, UT IP address is stored in the network, ouside of RAN, so does the pre-shared security related key.

For the RRC_IDLE UTs, the network has to broadcast the system information through its BCH (Broadcasting CHannel). The system information is needed by the UT and the paging messages.

In the tracking area where the cell is associated to, a unique ID should be assigned to the idle UT. After being assigned by a tracking area, the UT should wake up from time to time to listen to the paging channel if its relevant paging message arrives or not.

4.1.1.1 Paging

The paging procedure in WINNER (see Figure 4-2) follows this sequence:

- The GW-CP initiates the procedure by sending the paging message to each BS in the tracking area in which the UT is registered. The information concerning the UT's tracking area can be found in the mobility management (MM) context stored in the serving GW-CP of UT.
- The UT generates a paging response on NAS layer, and sends it back to the GW-CP based on the NAS level routing information.

The PICH and PCH in WINNER frame structure are shown in Figure 4-3. A UT is allowed to be active once per Discontinuous Reception (DRX) cycle, and the duration of DRX cycle can be either UT or network specific.. WINNER recommends that the network specifies the DRX cycle as multiple Paging Occasions. The paging occasion is multiple of the super frame duration. Both UT and network must know the DRX cycle since the PICH should be commoly known by the network and the UT in order to allow the UT check CPIH at every DRX cycle.

During each paging occasion, multiple paging indicators can be carried on the PICH which is realized by using L1/L2 signalling channel. A UT should check whether its paging group indicator bit is set. If so, it should read the Paging Channel (PCH), which shares resources with other control/user traffic. Precise UT identities inside paging messages are carried by PCH.

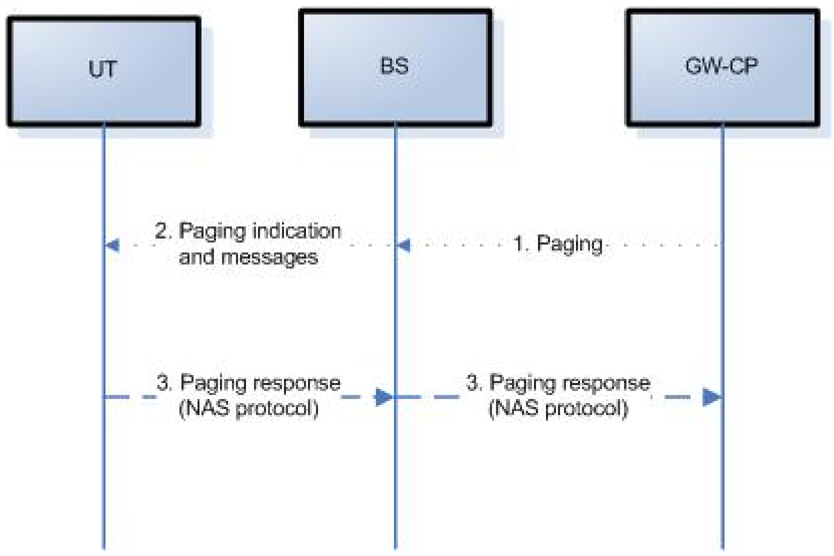

Figure 4-2: Paging procedure

A paging message shall be made of two fields:

- *Paging reason* which will be at least 3 bits, since at least 8 paging reasons should be supported.
- *Paging identity* which should be the Temporary Mobile Subscriber Identity (TMSI, 32 bits) of the UT.

As a result, each paging message will be at least as long as 35 bits.

In WINNER, the length of Resource Block (RB Chunk) for PCH (using BPSK with ½ rate) is 48 bits for the FDD mode, or 60 bits for the TDD mode. Therefore, one RB can only contain one Paging message. Since one paging message will occupy 0.6912ms, the paging capacity of WINNER equals to 1447 pages per second on each PCH channel.

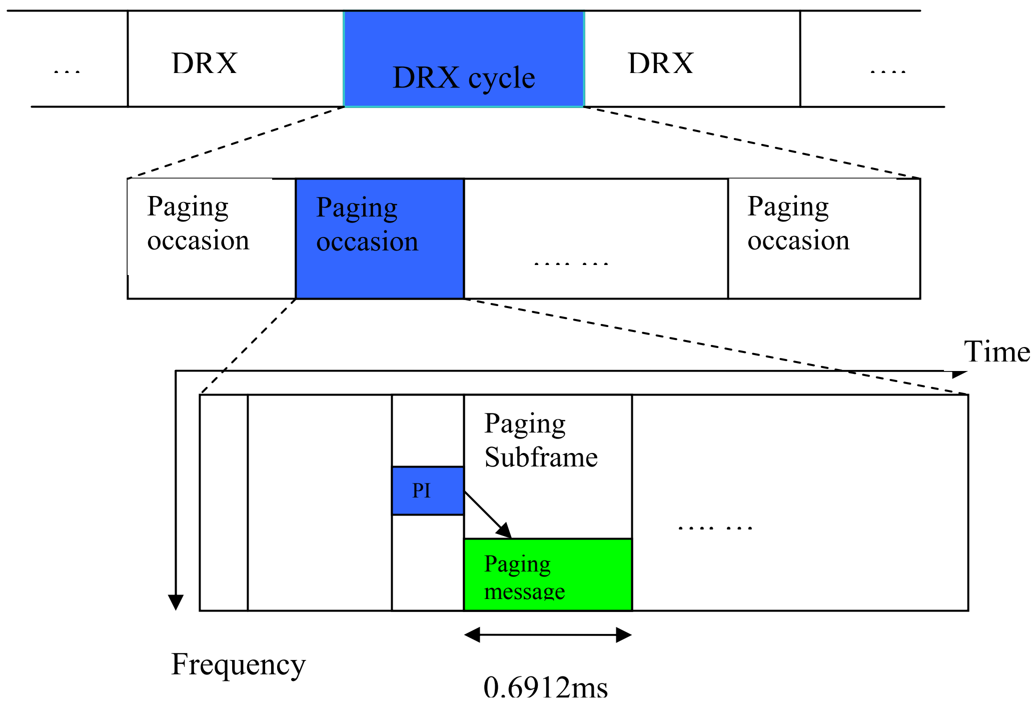

Figure 4-3: PICH and PCH in WINNER

4.1.1.2 Cell selection and reselection

As shown in Figure 4-4, from the detached mode to the idle mode, a UT performs a cell selection. Later, during the active period broadcasted by the network, the UT has the chance to reselect the cell. TA stands for Tracking Area

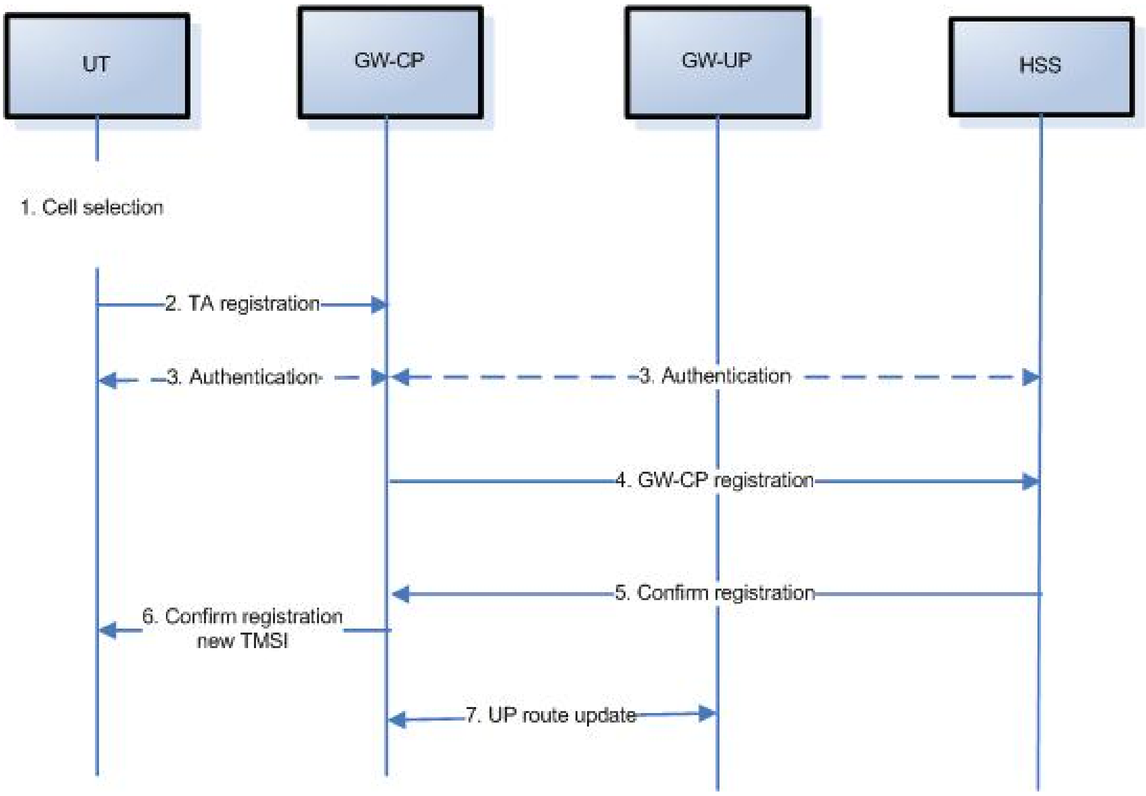

Figure 4-4: Cell selection procedure

Based on the BCH and CPICH (Common Pilot Channel) channels, the UT is able to camp in a cell, which shall be better against other candidates according to the cell selection/reselection criteria. Moreover, thanks to the provided HIS system, the UT velocity can also become one of the criteria for the cell selection/reselection. The cell reselection will be performed when the UT realises the ranking of the cells are reallocated.

The UT's tracking area registration can happen due to cell reselection, the expiration of TA update timer, initial registration of UT, or HIS triggered TA update. After the radio cell selection by the idle UT, the tracking area information may need to be updated to the UT's new GW-CP. UT sends its Temporary Mobile Subscription Identity (TMSI) previously assigned by the old GW-CP to the Mobility Management Entity in the new GW-CP when those GW-CPs are different. Based on the TMSI information, the old GW-CP forwards the UT context to the new GW-CP. After the authentication process between the UT and the new GW-CP, the new GW-CP registers the UT in the operator Home Subscriber Server (HSS). After the deletion of the UT context in the old GW-CP, the new GW-CP is confirmed by the HSS about the success of the registration. The new GW-CP issues then a new TMSI to the UT with updating the U-plane route. The procedure is shown in Figure 4-5

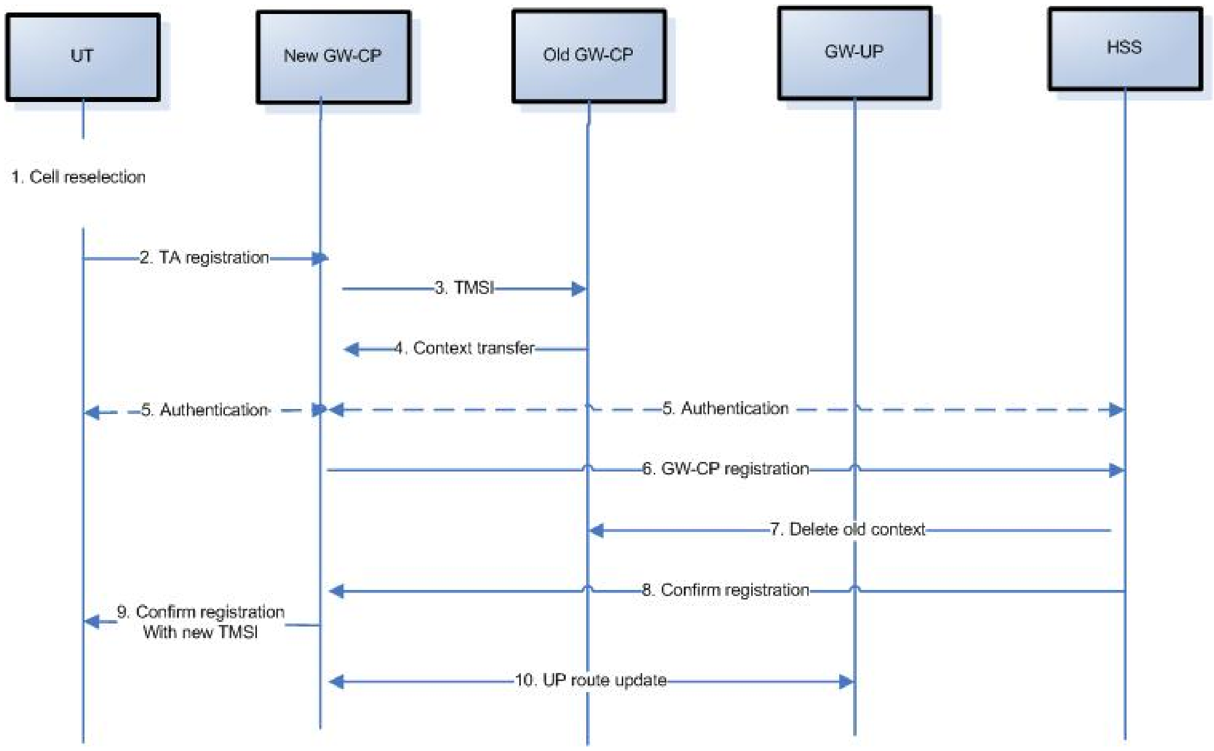

Figure 4-5: UT cell reselection procedure

4.1.1.3 Context transfer in idle mode

In idle mode, the UT is switched on but no connection is set up between the UT and the network, i.e no exchange of user data is possible in this mode. A procedure to update the user location is needed as a consequence of the UT's movement within the radio coverage to ensure that it can page the UT for an incoming call at any time. The location update can be done from the network side or from user side:

- Network side: Paging mechanism
- User side: Request for a location update

Module Name	Contents	Type	Kept in
User	User ID General Preferences Context by default	Static	GW_{CP}
Context	Context ID Context Restrictions User preferences	Static	GW_{CP}
Service	Service ID Suspension status Subscribed QoS	Static	GW_{CP}
Subscriber	Subscriber ID	Static	GW_{CP}
Service provider	Service provider ID	Static	GW_{CP}
Service profile	Service restrictions User availability User settings	Static	GW_{CP}
Terminal	Terminal category Hardware configuration Supported service list	Static	GW_{CP}
Access network	Network ID Static network configuration	Static	GW_{CP}

Table 4-1: Types of Static User Context

In accordance to location update, in the case where UT are moving to the target BS which is connected to different GP_{CP}, the so-called static user context that contains the user preferences must be transferred to the target GW_{CP}. Support of the user preferences will depend on the capabilities of the UT. If the capabilities change, the degree of support of the user preferences may change too. This is described in Table 4.1:

In this section, we consider two different scenarios for location update from the user side. The first one is the situation where UT moves to target target BS which is connected to the same GP_{CP}, thus the user context transfer is not needed, while the second one is the situation where UT moves to the target BS which is connected to different GW_{CP}.

Scenario 1: UT initiates to change the BS in the same GW_{CP}
UT mobility within the same GW_{CP} practically does not require exchange of information since all data is kept in the GW_{CP}, e.g. TMSI, authentication, etc. Therefore, there is no need of doing context transfer here. Furthermore, since the TMSI and the authentication are also kept in the GW_{CP}, the new TMSI and re-authentication are not necessary. Figure 4-6 shows the signalling flow of the UT initiated location update when it moves to a target BS within the same GW_{CP}. The UT sends a location update request to the target BS. After receiving this request, the target BS sends a User ID Check Request to the source BS to check if the UT is registered. After receiving an Ack from source BS which inform the validity of the UT, the target BS sends an update location message to HIS. Afterward, HIS sends a message to source BS to update its database. Finally target BS sends a new route advertisement to GW to update the routing table.

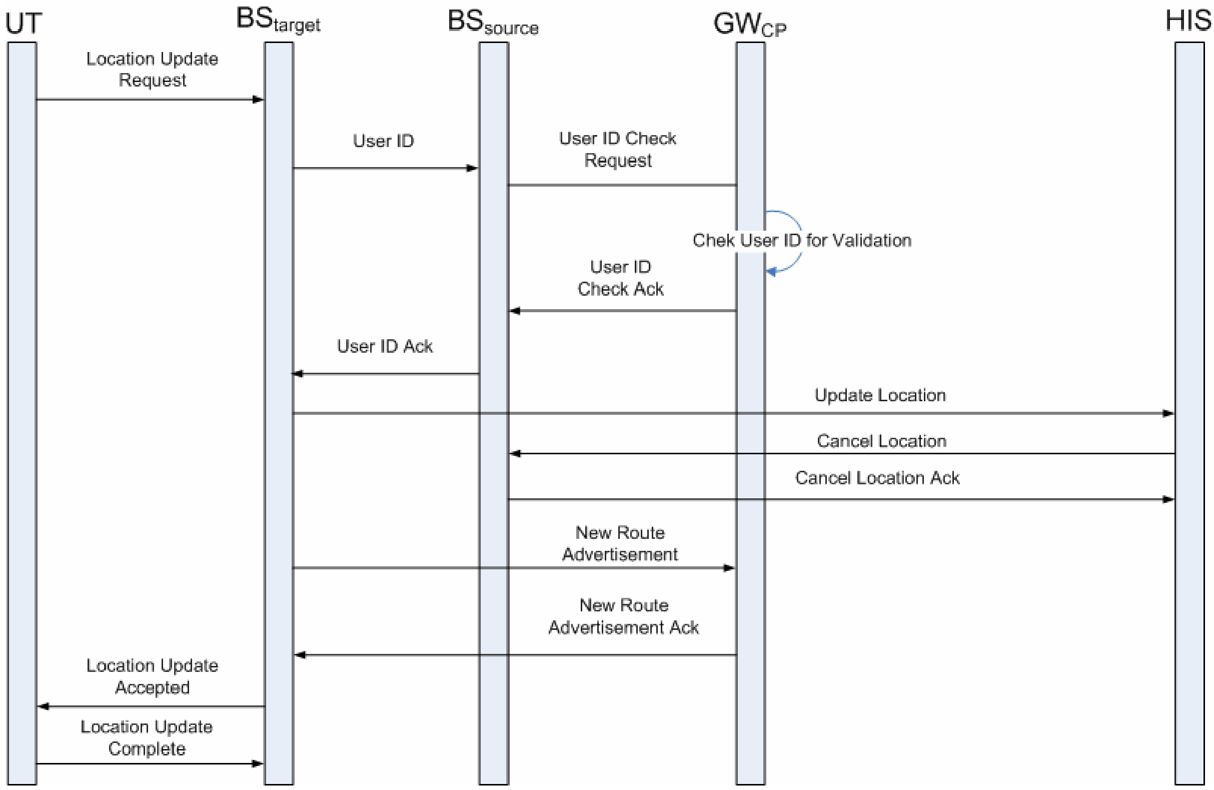

Figure 4-6: Signalling flow of the UT initiated location update within the same GW_{CP}

Scenario 2: UT initiates to change BS in different GW_{CP}
In this scenario, context transfer between the source GW_{CP} and target GW_{CP} is necessary since the target GW_{CP} needs to have all the information about the UT. Other than context transfer, it also involves the process of renewing TMSI and authentication. The new TMSI will be assigned to the UT by the target GW_{CP} and the authentication process will also happen in the target GW_{CP}. Figure 4-7 shows the signalling process during the UT location update within different GW_{CP}.

The flow diagram shows the message exchange during location update mechanism in case that the UT changes the cell that is served by other BS which attached to another GW_{CP}. In this scenario, the target BS has to request to the source BS to verify the validity of the UT. After validation is successful, the context transfer also is performed from the source GW_{CP} to the target GW_{CP}.

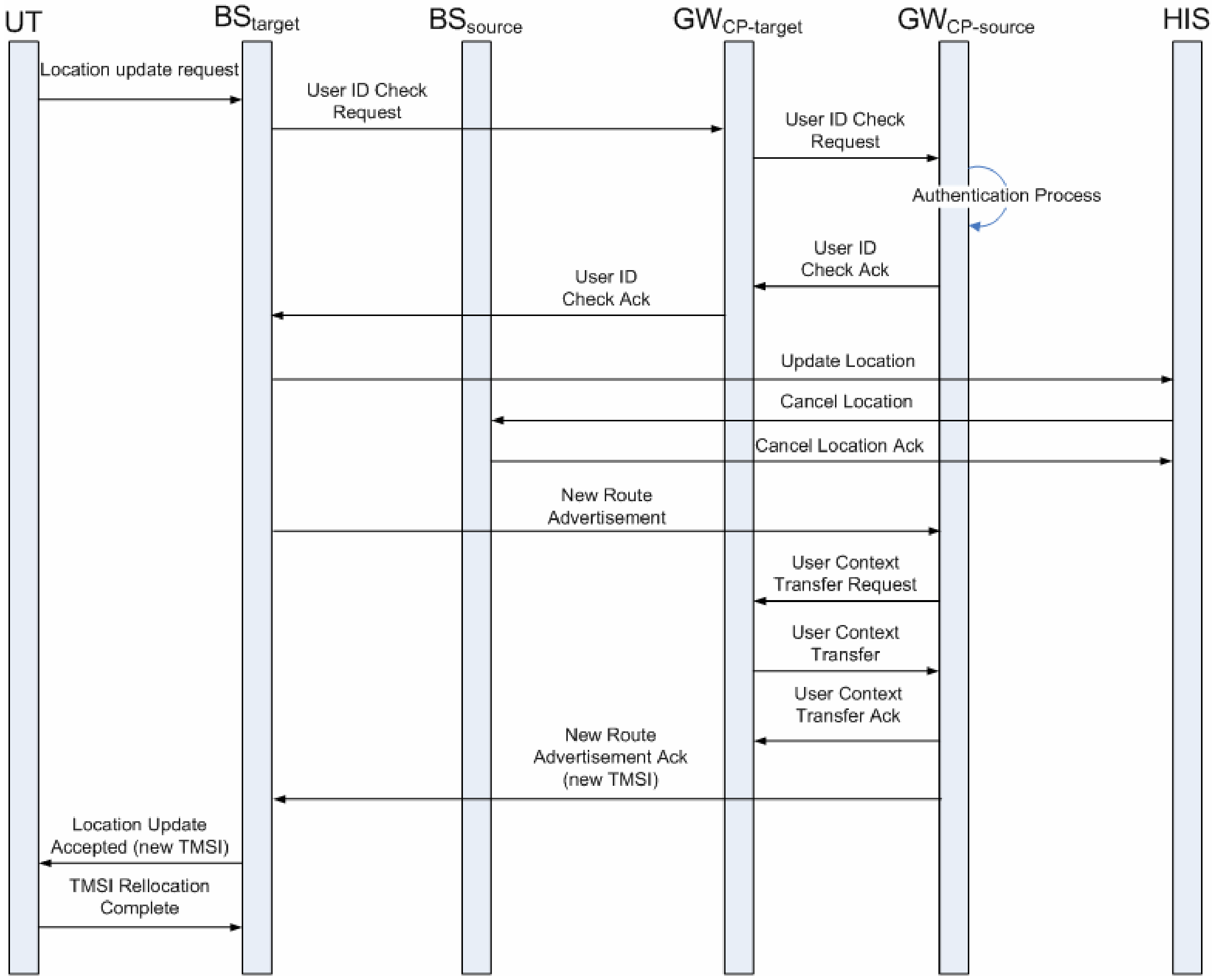

Figure 4-7: Signalling flow of the UT initiated location update within different GW_{CPs}

4.1.2 Broadcast of system information

WINNER is able to interact with legacy RANs through to be defined interfaces. Some specific details of the legacy RAN/RAT could impact directly on the Multi-mode UTs, that could support both WINNER and one or more specific legacy RATs. The WINNER Multi-mode UT, connecting to WINNER RAN, is able to decode the encapsulated information coming from a determined legacy system for inter-RAN handover and other functional purposes.

The WINNER RAN should be able to transport in a transparent mode (encapsulated) control messages coming from a legacy RAN to be transmitted to the UT.

To enable fast frequency scanning and cell registrations to the proper RATs for the Multi-mode UTs, the WINNER system is able to broadcast the selected system information from the legacy systems according to the operators' policy. Such system information can be the frequency allocation to the specific legacy RATs.
The list of information to be transmitted by the broadcast channel is the following:

- Stable information for UT initial access.
- Basic properties of the RACH
- Cell ID, operator ID, TX power mask, FDD/TDD duplex mode info.
- Cell-specific information that changes frequently.
- Shared band availability
- Spectrum sharing restriction parameters
- Pointer to next control channel (super-frame allocation table)
- Inter-system handover: Coupling point (gateway level versus IMS).

4.1.3 Two Stage Admission Control

WINNER system employs the load dependent multi-stage distributed admission control, which is characterized by the following features:

- The number of the entities involved in the end-to-end communication system are subject to the admission control, and the admission is not only decided by one entity
- The **'Token'** is assigned to the entity which/whose controlling domain has the most critical situation, e.g., lowest capacity
- The involving entities sends '**Flags**' with different tags (red, yellow and green) to others, and each tag has its own meanings
- The admission control is not immediately done when an entity receives a '**Green flag'** and a further checking of the most instantaneous situation is performed
- The rejection command is immediately sent when the '**Red flag'** is received
- A shared resource is re-partitioned when a '**Yellow flag'** (a soft-flag) is received '

Compared to conventional system, this protocol holds the following characteristics:

- The admission control is distributed but jointly developed by all involved entities
- There exists a load dependent sequence of decision polling, i.e., the ranking of the intermediate decisions is dynamic and depends on the load. Whose domain's load is highest, who receives the token to perform the admission check at the first step.
- The distributed decision is controlled by passing tokens between the involved entities and assigning flags to entities that immediately reject new admissions.
- The token is scalable due to service: Token for high data rate service may be located in GW, Token for real data service may be located in BS; even the network has the same rate. Such scalability is also applied to the DL and UL case.
- A soft-flag concept in relay enhanced cells for resource re-partitioning

In the following some examples are provided to explain the basic idea.

4.1.3.1 Gateway and BS scenario – Basic scenario

The initial simple / fundamental case is described first. In Figure 4-8, some basic network elements of the future network are shown, comprising the GW, BS, UT, and RN.

The GW provides the interface to the Internet and communicates with external routing functional entities. The GW also provides the anchor point for inter-RAN communication. It can be foreseen that in the future, GWs connect different RANs and the RANs will be considered as the elements of the whole communication networks. The BS performs almost all radio related functions for active terminals (i.e., terminals sending data) and is responsible for governing radio transmission to and reception from UT and RN in one or more cells. The BS is in control of relays (if used) and determines routes, forwards packets to the respective relay and takes care of flow control for the relays to ensure that they can forward the data to their associated UT.

The RN is equipped with relaying capabilities that is wirelessly connected to a BS, UT and/or another RN. As such it contains forwarding function and schedules packets on the radio interface. Furthermore system information broadcast, provided by the BS is relayed by the RN for an extension of the system coverage. One BS may communicate with UTs through multiple RNs, i.e., multi-hop.
As described in Figure 4-8, the admission control function is distributed among BS and GW. The BS takes care of the radio part during AC, while the GW takes care of congestion avoidance within the core network or other sub-networks.

Name the decision made by the BS as D1, and the decision performed by the GW as D2

Di = 0 in case of rejection; Di = 1 in case of acceptance, with i = 1 or 2.

The final decision for the incoming call will be a Boolean operation, i.e. D_final = D1 and D2.

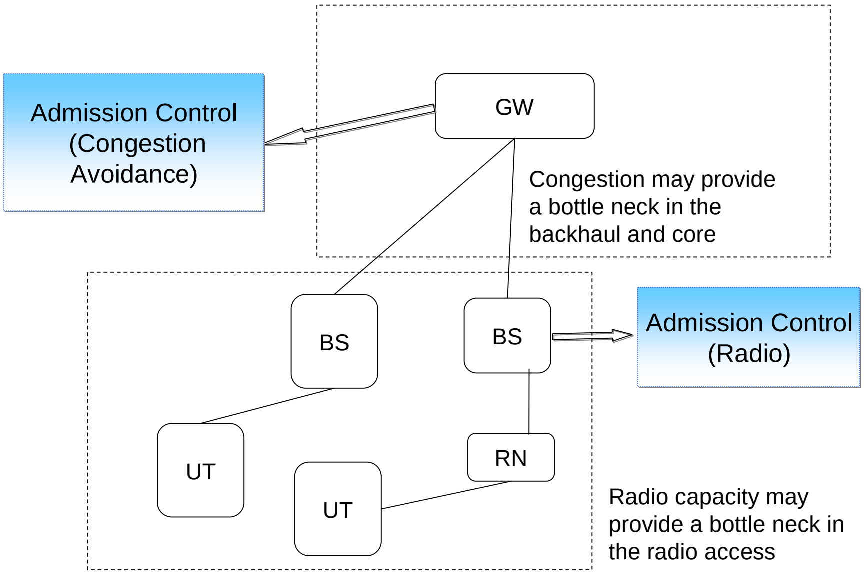

Figure 4-8: Admission control function distribution over GW and BS

4.1.3.2 Decision making sequence according to the network context

In this scenario, the involving entities where the admission control functions are distributed are ranked according to the system context. A typical example is shown in Figure 4-9.

In this scenario a token is introduced that determines the sequence of the AC. The token holder sets a flag in order to allow following-up procedures. As shown in Figure 4-9, the current load of the RAN is low, and the BS has relatively high capacity left. In the same time, the GW identifies a higher probability in congestion in the backbone or a limitation from other subnetworks which the expected traffic has to go through. In that case, the GW holds the token to perform the congestion prediction first before the radio admission control is performed.

In addition, the token assignment can be service dependent. E.g. with the same situation, incoming high rate data service (high FTP) needs different token assignment as voice like service. For instance, high traffic load demanding service requires an early check at the GW, low data rate real time service requires an early check at the BS.

For instance, if an incoming session is expected to add too much load in data rate in the backbone network or to other subnetworks, the GW will send the ‘red-flag’ to the corresponding BS where the UT is camped on. The BS then immediately rejects the session.

When the GW identifies the capacity in the backbone is sufficient, it issues a green-flag. After receiving the ‘green-flag’, BS checks if the available resources are sufficient for the incoming call, an ‘admission command’ is issued.

The RRM server and CoopRRM server shown in Figure 4-9 are the entities that provide enhanced information to the system, which help more intelligent ranking of the decision making sequence in order to assign the token. The RRM server coordinates radio resource for neighboring BSs, whereas the CoopRRM server coordinates the resources between future systems and legacy systems. In that case, the CoopRRM and RRMserver may serve as ‘Token Setters’, who assign Token to the right entity.

The sequence can be re-ordered according to the system context by more suitable token assignment.

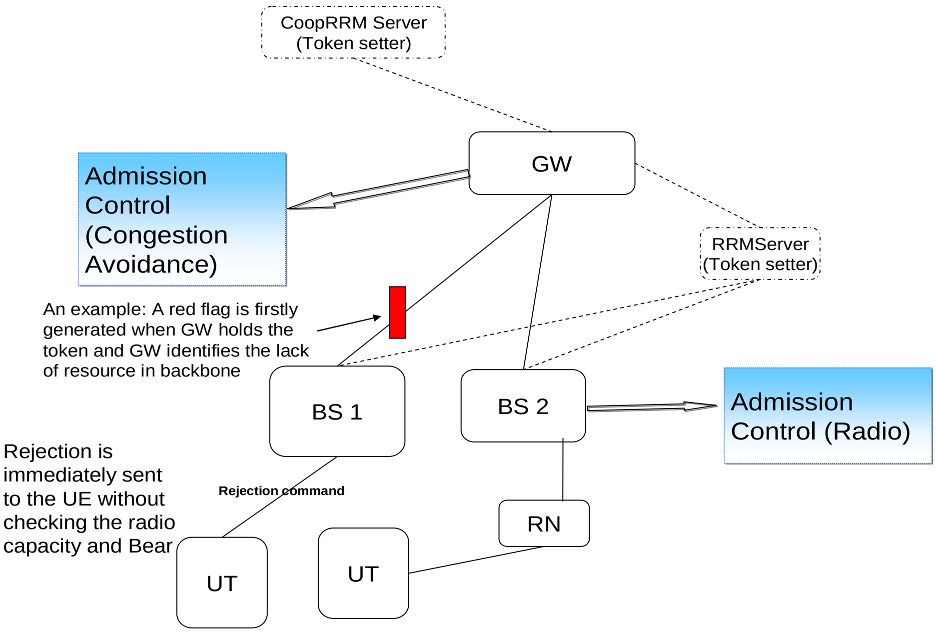

Figure 4-9: Token setting for sequential flag

4.1.3.3 Multi-hop and relayed network

The system with multiple relay nodes (RNs) supporting coverage extension and throughput enhancement also needs such protocol. The BS can not perform admission control only based on its available resource or only based on the available resources of the RN that should directly serve the UT. Also the intermediate relaying capacity between RNs and RN and BS has to be taken into account. In this case, the distributed mechanisms as explained for the GW and BS can be applied to the BS and RN.

The available capacities of all segments of all possible routing paths between the UT and the BS have to be checked before an admission command is generated. The segments are defined according to the direct connections among the RNs and the BS. The token holder is given w.r.t. DL/UL and the final relay node to the to-be-admitted UT. For FDD system, the capacities of UL and DL will be different, which makes the token holder assignment also different. For instance, as shown in Figure 4-10, the poorest link in the UL from the RN2 respect is S12U, therefore the token is assigned to RN2; however, in the DL, the poorest link is S01D respective to RN2, therefore the token is assigned to BS.

As an example, when RN_2 identifies a lack of resource for the incoming calls (bottleneck identified), it sends immediately a yellow-flag (soft-flag) with its marker/header to the central resource control unit in the cell (typically the BS) as step 2 depicted in Figure 4-11. The BS checks with GW about the core network resource and repartitions the resource and at the same time confirms to the RN_2 by sending the green flag. RN2 then admits the UT when it confirms the sufficiency of the resource.

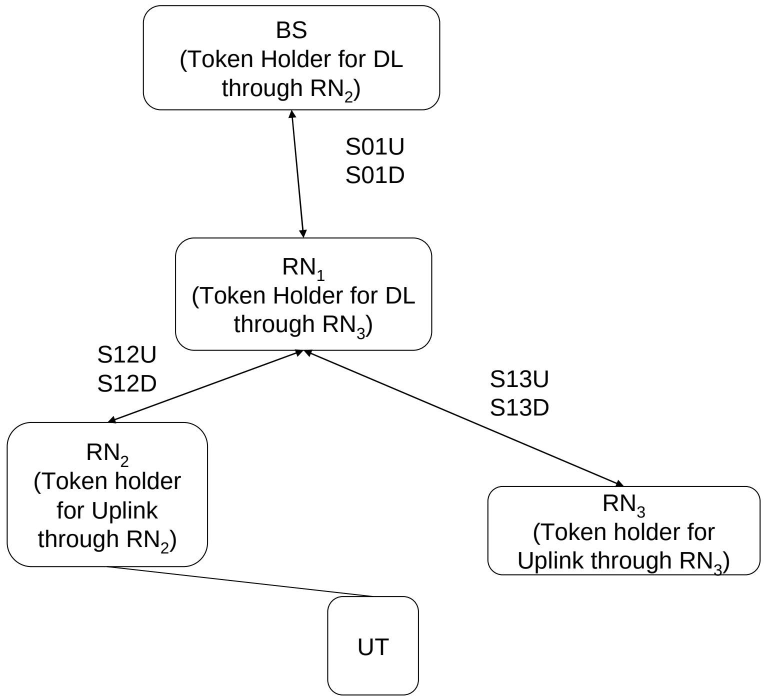

Figure 4-10: Segments and Token Holder

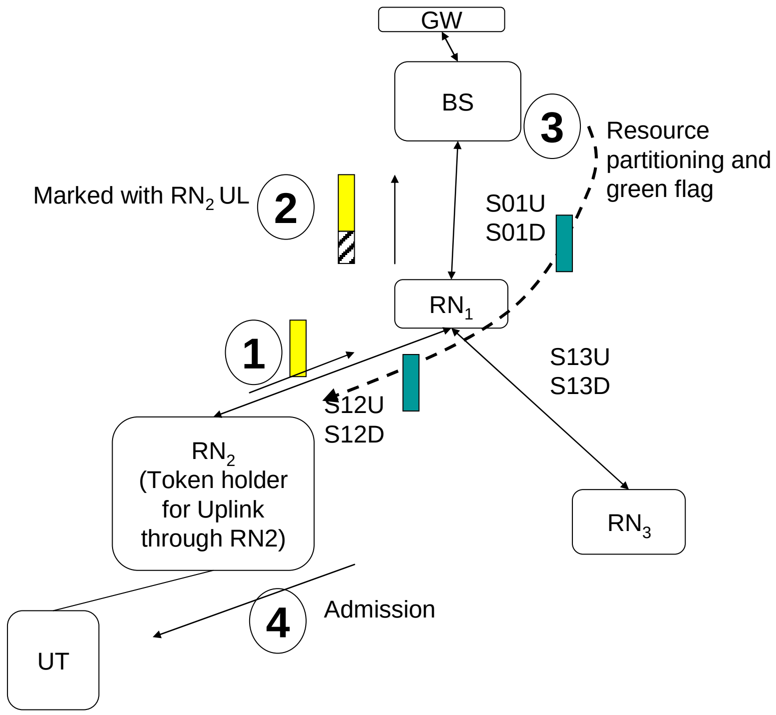

Figure 4-11: An example of the soft-flag and piggy-packed green flag along with resource partitioning

This process is different to the BS-GW case, since the resources among the involving entities are shared. The rational behind is that there is a potential trunking gain may be exploited (see Figure 4-12).

For each RN, it has to restore the optimal routing paths for any potential incoming sessions. Throughout the path, the bottleneck has to be identified w.r.t. the QoS expected from the incoming calls.

For systems with multiple segments, comprised by the sub-networks and relays, the number of stages can be extended higher than two, which however does not lose the generality and principle of this aproach.

The red or green flag goes always to the next decision maker. However, the yellow-flag (soft-flag) is among the coupled entities that may perform a resource repartitioning in order to allow the incoming calls.

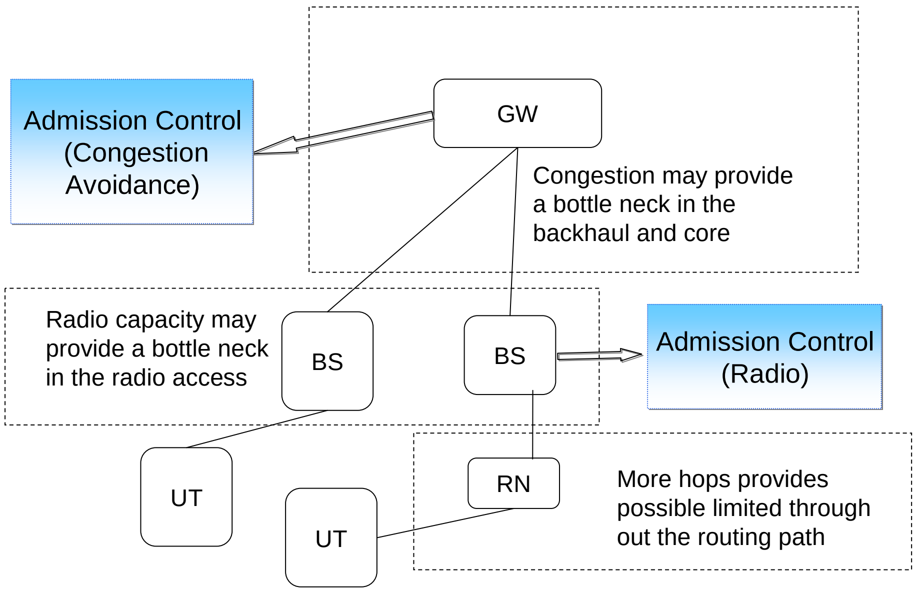

Figure 4-12: Relay enhanced network supporting Multi-hops

4.1.3.4 Other controlling functions

The default token can be set by the O&M during the network planning and the maintenance phase. When the load and capacity changes in the system, default token can be reallocated to the most severe entity following the principles/mechanisms described in the early sections.

4.1.3.5 Advantages of the approach

The two-stage Admission control has the following advantages:

- High overall system performance in terms of QoS (less congestion) and GoS (less dropping rate)
- Balanced decision load of the involving entities. The decision load is defined as the admission request intensity per time unit. Due to the balance, each entity has a lower load; therefore a potential decrease of the response time of the network entity may be obtained. (for instance, 10 request per second in classic solution can be reduced to 5 requests per second to one entity)
- No hectic inter-GW BS context transfer. In case of the GW is the limiting factor, a simple admission control only in the BS will normally result in a biased/wrong decision. Due to a wrong admission, the UT context has to be re-allocated from the GW to the BS. after dropping, the flow-context has to be completely deleted from the BS and user NAS context has to be reallocated back to the GW), this will especially release the burden of the future Iu-Flex (S1-Flex) interface
- Potential reduction of the air interface signaling load (RRC messages): since if the GW has already decided a rejection, the BS does not need to check the RAN capacity. That process sometimes needs to trigger measurement reports. For network triggered call, if the GW rejects the call, the paging channel, AICH, layer 1, 2 signaling are saved.
- No negotiation of BS-GW signaling needed, since GW holds the QoS information for Idle UTs and the policy enforcements information
- Much faster AC decisions than in conventional systems: (reason 1: See the second advantage; reason 2: in case of rejection, only a rot flag from the Token holder can reject the call immediately)

4.1.4 Establishment maintenance and release of a RRC connection

UT is switched on, scans available cells and connects with the most adequate. After camping the UT is able to sent and receive system and cell information, transmitted by the broadcast control channel (BCCH). It stays in idle mode until it sends a RRC petition of connection, when it is granted it switches to connected mode.

Figure 4-13 despicts the WINNER physical, transport and logical channels described in [WIND6.13.14]

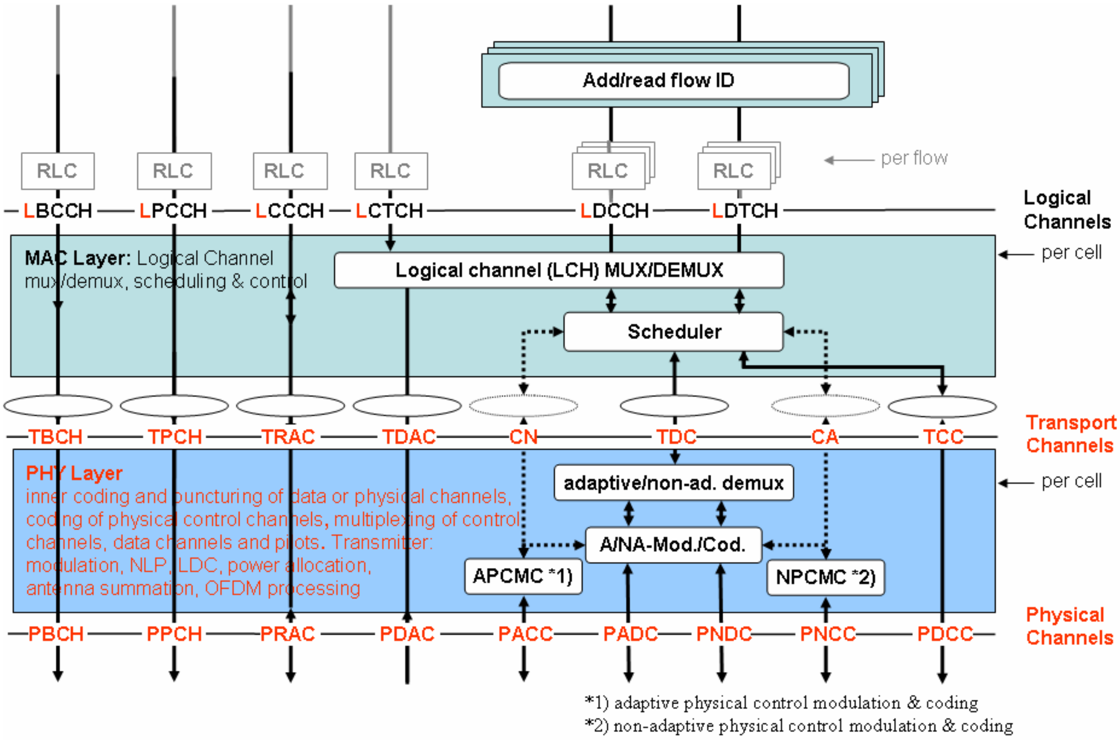

Figure 4-13: Physical, Transport and Logical channels

The RRC connection is defined as a point-to-point bi-directional connection between RRC peer entities in the UT and the WINNER RAN (BS or RN) characterised by the allocation of a Network Temporal Identity (NTI). At the end of the process the UT will have assigned a channel, usually a dedicated channel. An UT only can have only one RRC connection

The procedure to setup the establishment of a RRC connection, based on the UT initial connection in the uplink, is the following (see Figure 4-14)

1. The UT receives system information from the broadcast channel (PBCH), including PRAC, PPCH radio resource allocation and control information, necessary to be connected to the network.

2. The UT initiates set-up of an RRC connection by sending **RRC Connection request** message on the shared Randonm Access Channel (PRAC). The UT uses the PRAC because it has not having radio resource control connection with the network.

   Parameters: Initial UT Identity, Establishment cause.

   The message RRC connection request will be used in many circumstances, apart from of the initial connection (registration), including **maintenance and release of the RRC connection,** which are gathered in the access reason:

   - Registration (initial connection), Detach
   - Initiation of signalling (typically when switching from idle mode), signalling termination
   - Establishment of data traffic (with identification of WINNER data flow type)
   - Terminating of data traffic (with identification of WINNER data flow type)
   - Re-definition of WINNER type of data flow

- Terminating – cause unknown (e.g. low battery)
- Inter-RAT cell re-selection
- Emergency call
- In the case of messages related with establishment or release of data traffic, and also with other types of traffic, further signalling between the BS and GW and GW to HSS will be needed.
  In the case of messages related with establishment or release of data traffic, and also with other types of traffic, further signalling between the BS and GW and GW to HSS will be needed.

  Depending on the establishment cause, this message will progrees toward the GW (e.g establishment of data flow) and/or the HSS (e.g registration) to initiate the signalling processes is going to be initiated.

3. The Serving BS sends a **RRC connection setup** in the downlink PRAC, that repeats the initial UT identity to address only one sender UT (several UT can have sent a RRC Connection Request message at the same time on the common channel) . This message indicates the channel that has been allocated for the RRC signalling exchange. The BS could assign the common channel PRAC, but usually it will assign a dedicated channel for signalling (LDCCH) or for user data (LDTCH).

   Parameters: Initial UT Identity (that could be IMSI –included in the SIM- or TMSI), chunks allocation for the dedicated (LDTCH/ LDCCH) or shared (PRAC) channel.

4. The UT sends the message **RRC Connection Complete,** using the channel indicated in the previous command, it can be a dedicated channel for user data (LDTCH) or for signalling (LDCCH), but also for short messages could assign the common channel PRAC. Using this information the UT sends relevant parameters

Parameters: Integrity information, ciphering information, UT capabilities.

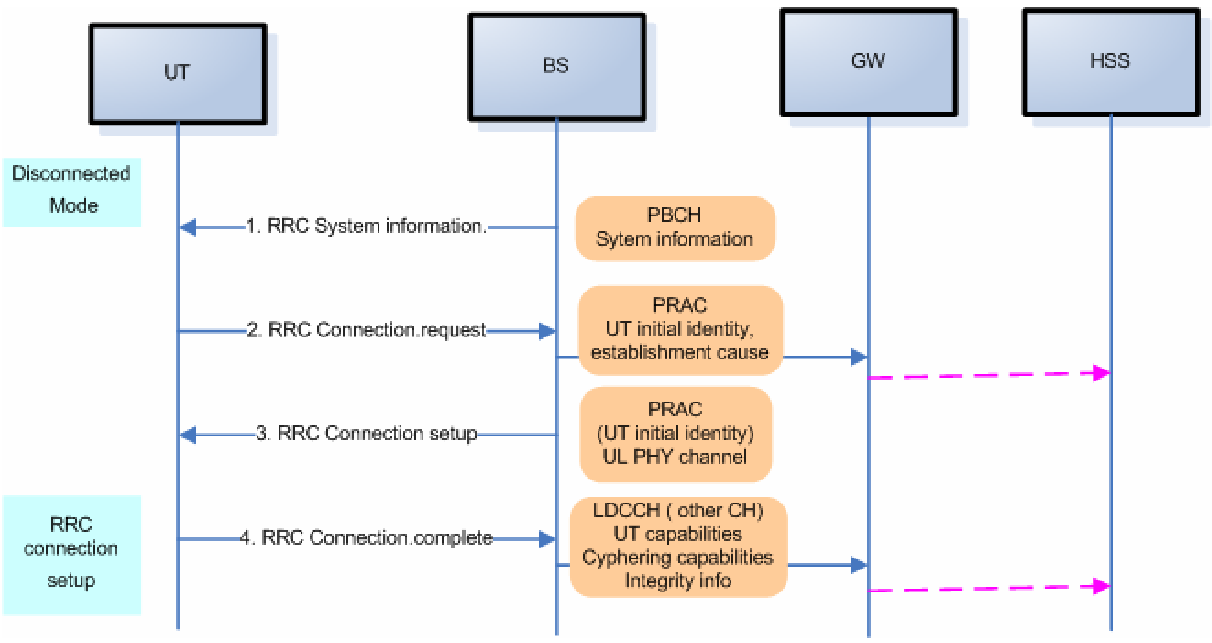

Figure 4-14: RRC connection process

4.1.4.1 Process after the establishment of a RRC connection

In this section the procedure followed after the initial RRC connection is drafted.
After the establishment of the RRC connection the process to obtain a complete connection is depicted in Figure 4-15. During this part of the connection establishment basically there is an exchange of NAS messages (initially not ciphered) with the NAS peer entities located at the GW and the core network, specifically the Home Subscriber System (HSS), Considering that NAS layer are not totally in the scope of this section, only is considered its interaction with lower layers is given.

NAS messages are transported in transparent or semi-transparent mode by the RRC layer. In transparent mode the GW do not decode the message, and in semi-transparent mode, it forward it to the NAS entity, but also it decode the message, a flag inside of the message distinguish this two types of messages. The NAS messages are transported using the messages RRC UL Transport (direction UT towards the HIS) and RRC DL Transport (direction UT towards the HIS).

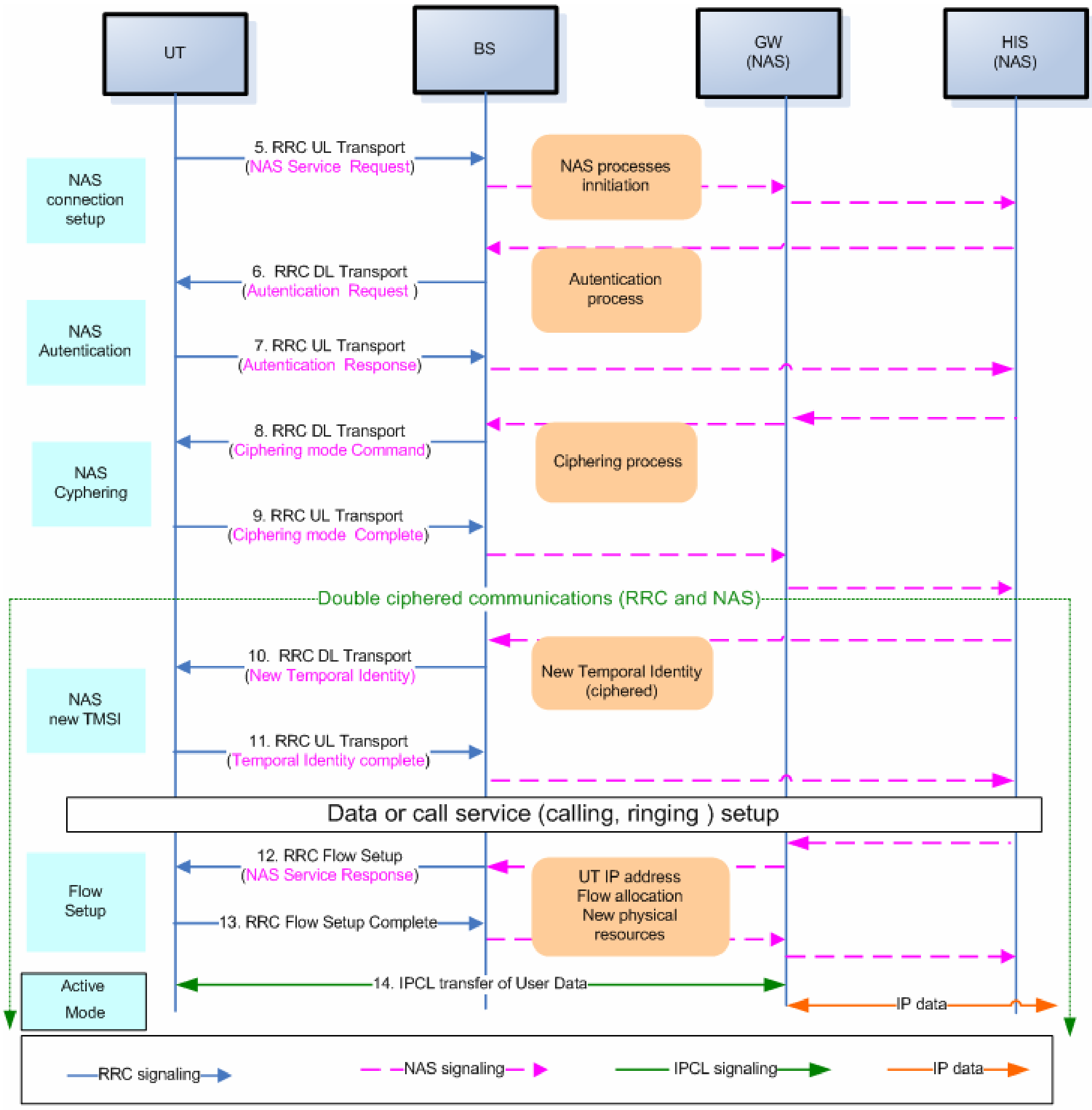

Figure 4-15: Processes after initial RRC connection stablishment (NAS connection stablishment and flow setup)

5. NAS Service request, The UT sends this message, to establish a NAS connection with a peer entity in the core network (the NAS layer in the HSS). The first steps are the UT Authentication and the Ciphering setup. This message uses the RRC transport service that included an embedded NAS-service request message.

   Parameters: NAS messages and HSS indicator

6. Authentication request, embedded on a RRC DL Transport message. As consequence of the previous message, the NAS send an authentication request message that is transported using a RRC transport message. The authentication is carried out using a private key stored in the UT (SIM card) and a corresponding public key in the HSS, using this public key the HSS send a coded message (challenge) that the UT have to decode using the private key
   Parameters: Coded message using the public key stored in the NAS.

7. Authentication response, embeddede on a RRC UL Transport message. UT answer to the HSS coded message
   Parameters: Decoded message using the UT private key.

8. Ciphering mode command. Initiation of ciphering process. There is a double ciphering process, one at RRC level, covering the radio communications; in this case the terminating nodes are the UT and BS, and other at NAS level, between the UT and the GW (and HSS), this last ciphering is also used by the User plane transmission (IPCL layer). Ciphering at RRC level is described in section 4.1.11 and ciphering at NAS (and IPCL) is described in section 3.8
9. Ciphering mode complete. After this message the double ciphering is established.

10. New Temporal identity. The HSS sends a new ciphered temporal identity (TMSI). The previous initial temporal identity used on RRC Connection Request was send without ciphering by radio, therefore it is not protected.

11. Temporal identity complete. The UT acknowledges the new temporal identity

After the previous message the network or UT, is able to ask for a data service (using a RRC connection setup message in which the necessary flow types can be requested), or a call (using a ringing protocol, not described here), in a safe maner.

12. RRC flow setup, including a NAS message NAS service response. Using this message the network gives the necessary information to UT to receive or transmit the coming information

    Parameters: IP address (provided by the GW), Physical resources of a dedicated or shared channel, flow allocation.

13. RRC radio flow setup complete. The UT acknowledges the assigned resources.

14. IPCL transfer of user data. After the previous message it is establish a IPCL channel between the UT and the GW, in which it is established a header compression protocol, an a IPCL service of user data transfer between IPCL entities (see sections 3.2 and 3.3)

4.1.5 Active mode micro-mobility

4.1.5.1 Intramode handover

Intramode handover is the handover between Radio Access Points (BSs and RNS) operating in the same mode, there are three possibilities: intra BSs, intra RNs and between RN and BSs of the same mode. This type of handover includes the intra-cell handover where the user remains in the same mode (an example, the change of frequency in the same cell) and the inter-cell handover between cells of the same mode. The basic trigger for inter-cell handover is the received signal strength, but also, the load of the neighbour cells, congestion situations, increased interference, the location of the user, etc. The intra-mode handover (between RNs and/or BSs) for example could be triggered when the received signal strength is below a fixed specified minimum value.

In the WINNER system, a user terminal may be in the coverage area of several cells of the different WINNER modes, so the number of cells which the terminal will use to perform measurements to them could be very large and cause limitations to the handover process. For the intramode handover the number of cells for the user terminal to measure could not be very large, but the use of neighbouring cell lists could improve very much the handover process.

The main problems concerning the number of cells to measure are:

- If measurement duration is fixed, then when the number of cells to measure increases, less samples are measured for each cell. The accuracy of measurements consequently depends on the number of cells to measure.
- If the number of samples necessary for obtaining the average value is fixed, then measurement duration will increase with the number of cells to measure. Besides, measurements will not be performed in a limited amount of time for all cells, and thus measurement values will not be comparable.

Even if the preferred target cell identity is specified by an independent decisional entity (depending in this case mostly on signal quality, or cell load, or location information), before performing handover, it will be necessary to have information on:

- The signal strength or quality on that preferred target cell, at the mobile terminal location;
- The load of the cells with highest (or sufficient) signal strength or quality on that preferred mode.

Consequently, it is necessary that the mobile terminal performs signal strength, signal quality or/and load measurements on some cells of the preferred target mode.

In order to restrict the number of cells to measure, the mobile terminal can be supplied with a list of the neighbouring cells, that are the most likely to fulfil the signal strength or quality requirements at the mobile terminal location.

The way neighbour cell lists are sent to the mobile terminal shall be optimized, depending on:

- The mobile terminal measurement capacity, which depends on the physical layer of the system the mobile terminal is currently camping on or connected to;
- The number of possible target cells of that specific mode;
- The presence of a controller entity in the current system;
- The accepted signalling load on the air interface;
- The possibility, for the mobile terminal, to receive dedicated messages containing neighbouring cells information.

Thanks to the use of neighbouring cell lists and of the measurement optimization process associated to it, measurements will be restricted to the cells useful for intramode handover

The criteria to choose the users (and their number) that perform the handover must be determined: it could be services based criteria (Speech users are handed over first, then another service, etc.) or resources based criteria (the users consuming lot of resources are transferred first, etc.) or users based criteria (bronze users are handed over first, then silver users, then gold users, etc.). The intermode handover algorithm in the UT performs periodic measurements and, when a trigger for intramode handover is activated, then the signals included in the neighboring cells list is measured and ranked as better candidates to handover. Then the first target cell is selected and the Admission Control decides to accept the user in the new cell or not. If the admission decision is positive the user performs the handover. Otherwise, it selects another cell from the target list and so on, until either the user is accepted in a cell or the user is rejected.

In Figure 4-16 the intramode handover RRC signaling is presented. The trigger for an intramode handover could be low signal strength at the user terminal, change of location using information from the HIS, increased interference trigger in the BS, cell congestion, etc. When there is an intramode handover trigger, the current BS gets the list of the neighbouring cells from the HIS and sends it to the UT. Now the UT knows which cells to measure the signal strength and sends the measurements back to the BScur (current BS). The BScur then makes a list of the possible target cells and sends the handover request to the new BStarg (BStarget1). Then the Admission Control on that BS is activated. If the admission is rejected then the BScur sends the HO request to the next target BS (BStarg2). When the Admission Control accepts the handover, the BStarg sends the HO request acceptance message to the BScur, which sends the HO command to the UT. Then the UT sends the HO completed message to the target BS to request a radio resources and the BStarg response and sends also a HO completed message to the BScur to release the radio resources of the UT. The BScur acknowledges and releases the radio resources and the handover is completed.

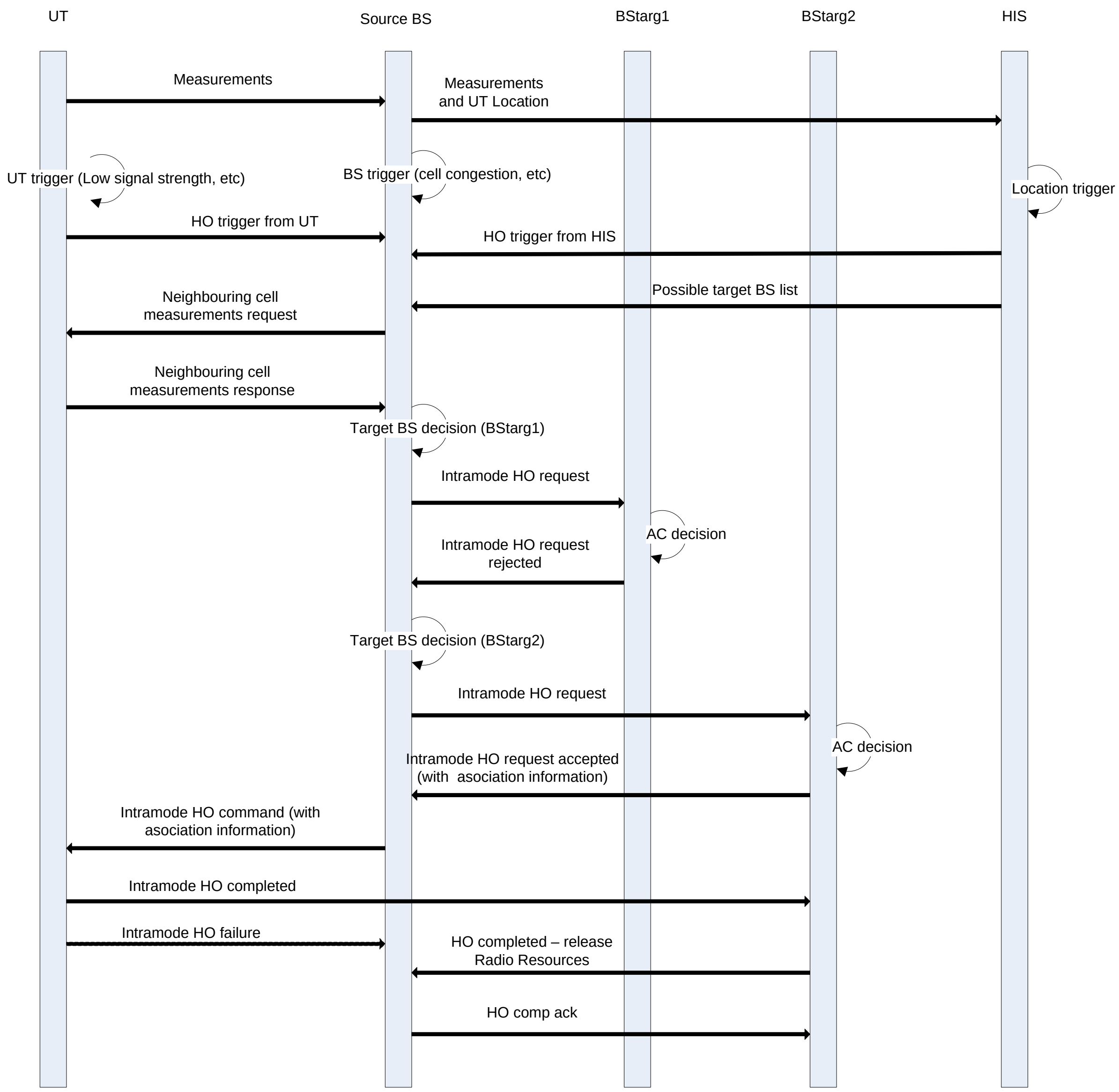

Figure 4-16: Intramode handover signaling

4.1.5.2 Intermode handover

Intermode handover is defined as the switching process between two cells of different WINNER operational modes with different cell size. The typical WINNER scenario is expected to be the one where the cells of the different modes are overlapping either completely or partially.
The triggers for intermode handover are:

- Insufficient signal strength or signal quality,
- Congestion in one mode,
- User mobility,
- Terminal location,
- Other physical layer based triggers and algorithm based triggers.
- Also many other triggers can initiate an intermode handover, such as user's location, user's received service requirements (i.e. datarate), etc.

In WINNER, three operational modes have been defined, namely the wide area, metropolitan area and the local area, with the aim to have a scalable and optimized system always offering the best network performance. It is assumed that the higher mobility support corresponds to wide area mode, after the metropolitan area and the lower to local area. The higher datarate corresponds to the local area, after the metropolitan area and the lower to wide area.

The handover from wide area to local area could be initiated either by the UT or the BSwa. However, the decision for the handover acceptance will be made by the Bswa/ma. The specific triggers that could initiate the handover from a wide area cell (usually using FDD PHY layer mode) to a local area cell (usually using TDD PHY layer mode) are the following:

- Need for higher datarate provided by the local area cell (UT or BSwa initiated)
- Congestion in the wide area cell based on cell capacity and load (BSwa initiated)

In the first of the process, the UT will check whether its velocity is lower than a maximum limit, this information will be derived from the location data contained in the HIS reports and therefore whether it can switch to the local area mode. Then, the UT will send a request for handover to the BSla that will be transmitted to the BSwa. The BSwa using the information provided by the BSla within its cell, will check whether the load of the target local area cell is low enough so as to accept or decline/queue the handover request. In case of a handover acceptance, the BSwa will also send to the UT the needed information on the target and the neighbouring cells, enabling the UT to perform measurements and complete the handover.

The handover from local area to wide area could be initiated either by the UT or the BSla. However, the decision for the handover acceptance will be made by the BSwa. Examples of triggers are:

- Increase of terminal velocity (UT initiated)
- Loss of local area coverage (UT initiated)
- Congestion in the local area cell, use of QoS restrictions and user priorities could be used for deciding which users should handover to the wide are cell (BSla initiated)

In order to obtain a fast and efficient handover from wide area to local area, we introduce a pre-trigger status in which the needed measurement for handover are requested and provided to the involved logical nodes (BSla, BSwa, UT). The UT will request the approximated inter-mode measurements from the BSla after the activation of a pre-trigger. A pre-trigger can be either an algorithm trigger that indicates the possibility of a handover (requested service not available in the local area cell), a request from BSla to handover (due to reaching congestion limits within the cell) or a PHY trigger based on the UT intra-mode measurements (e.g. BER>pre_trigger_limit initiates the exchange of measurements with the BSla while BER>trigger_limit>pre_trigger_limit necessitates the handover to wide area). After the pre-trigger activation, the UT will send its intra-mode measurements to BSla and request the approximate measurements for the Bswa. However, it won’t use them to request a handover unless a trigger that necessitates handover will be activated. After acquiring the wide area cell measurements the UT will send a handover request to the BSwa indicating the reason for this request. Based on this information, as well as on the BSwa cell and BSla cell information, the BSwa will decide to accept, decline or queue the UT handover request.

When there is an intermode handover trigger, the algorithm tries to find the best suitable mode for the user to handover to. This decision is based on several criteria which were analyzed above and also on the trigger that requested the handover. The target modes (if there is more than one suitable) are listed and ordered by preference according to the above criteria. For the selected mode a list of target cells is created and it is checked with the admission control to which cell the user can be admitted.

The RRC signaling process for an intermode handover from local area to metropolitan area (coordinated by a BS of the wide area) is presented in Figure 4-17.

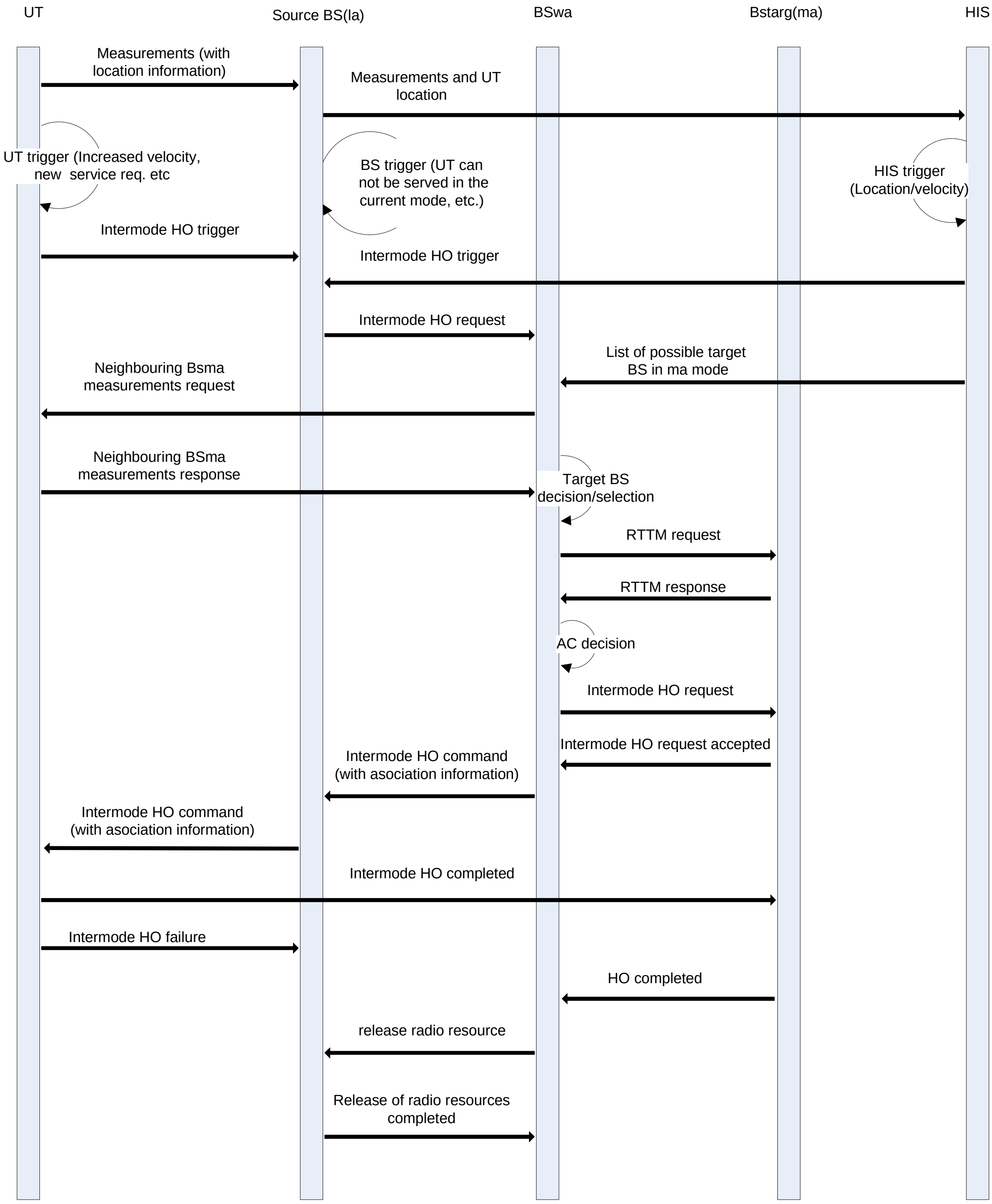

Figure 4-17: Intermode handover signalling process from LA to MA coordinated by BSwa

In this process, the BSwa is the base station that coordinates the handover procedure. The BSwa receives the request from the current BS for an intermode handover, gets the measurements from the mobile terminal and the BScur and decides the list of modes and BS of each mode that are suitable for the user. Then all the messages are exchanged via the BSwa in order to complete the handover.

4.1.5.3 UT measurement reporting

The RRC determine which radio resources that are available using the measurements done by the UT lower layers (but also at the BS and RN). After the measurement condictions have been achieved, the measurement is taken, and then it is transmited the measurements reports from the UT RRC to the RRC of the node element that is controlling the UT radio resources. In distributed RRM situations this entity is the BS RRC, in centralized RRM this entity is the optional RRM server and in cases of inter-system RRM the entity is the CoopRRM).

The RRC allows to the MAC to assign radio resource to the different users. The measurements are configured and transmitted using the control SAPs. Figure 4-18 shows this principle.

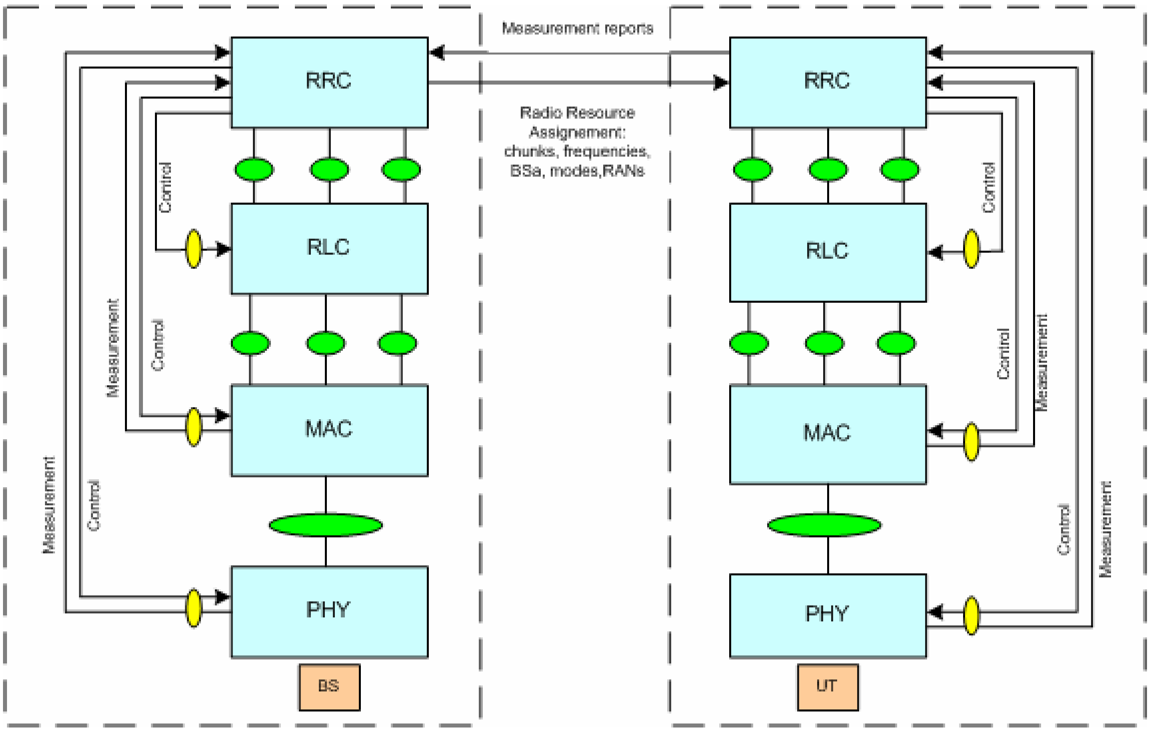

Figure 4-18: Measurements and RRC assignement of radio resources

Measurements to be performed by an UT (and also other physical node) can be configured using dedicated control messages; these measurements can be classified in four categories:

- Intra-frequency. On the same frequency in which the UT is operating
- Inter-frequency. On other frequency band
- Intersystem. On other networks
- Location determination. Related with UT positioning
- Reports. Measurements not related with physical radio channels (e.g BLER, BS load)

In RRC IDLE state, a UT shall follow the measurement parameters defined for cell reselection specified by the broadcast messages. In RRC_CONNECTED state, a UT shall follow the measurement configurations specified by RRC directed from the RAN in the same way is done the currently deployed systems.
Figure 4-19 present the proposed protocol, The UT (but also BS and RN) after been configured wait to be triggered by the specified conditions, then take the requested measurement and sent the results to serving BS (but also other nodes could be the final destination of the measurement as the RRM server or the CoopRRM entity.
The measurement_config command can:

- Setup. Setup a new measurement
- Modify. Change the current measurement setup
- Release. Stop a measurement (typically a periodic measurement)

Measurements can be configured as periodic but in some cases could be even triggered. Also the determination of the time advance between current and target base of time (e.g. the measurements of pilot channel of the target BS)
The use of neighbouring cell list will improve mobility management, reducing delays introduced by system functions as handover. Assuming a reuse of frequency of 1 (but there could be intra-cell frequencies split, for interference reduction reason), the predominant measurement will be intra-frequency measurements

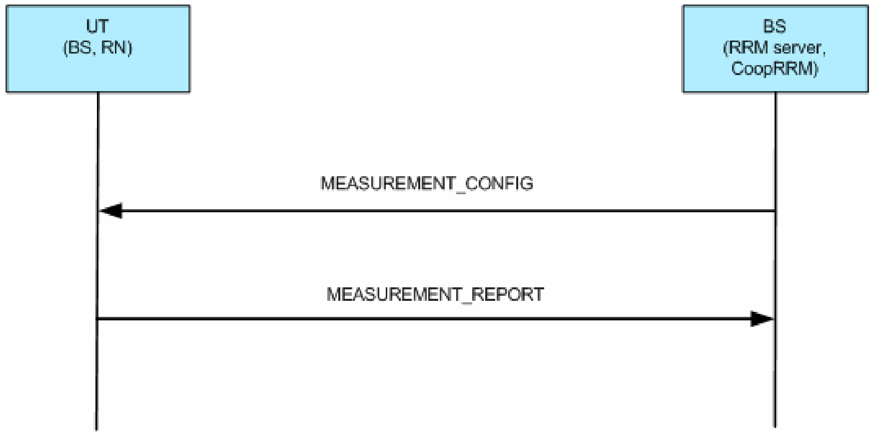

Figure 4-19: Measurements control and reporting

Depending on whether the UE needs transmission/reception gaps in the OFDMA superframe to perform the relevant measurements, measurements are classified as gap assisted or non gap assisted. A non gap assisted measurement is a measurement on a cell that does not require transmission/reception gaps to allow the measurement to be performed (e.g inter-frequency measurements). A gap assisted measurement is a measurement on a cell that does require transmission/reception gaps to allow the measurement to be performed (typically in the same band). Whether a measurement is non gap assisted or gap assisted depends on the UT's capability and current operating frequency. The UE determines whether a particular cell measurement needs to be performed in a transmission/reception gap and the scheduler needs to know whether gaps are needed

Measurements are essential inputs for the RRC functions, measurements affect to the design of the WINNER PHY, MAC, RLC and RRC from the configuration and physical procurement of these measurements, to the transport of these measurements to the logical entities that need this information, including the definition of the protocol for the transport this information.
In particular, the WINNER system should at least provide to the BS and to the optional RRM server (or/and cooperative RRM) entity a set of measurements for handover and other RRC functionalities:

- **Received signal strength (RSS)**, **Interference level and Signal to Interferenece and Noise Ration (SINR) ratio.** This must allow concluding on the reception quality of the actual configuration and the possibility (or the necessity) of doing a handover to other cell or radio access technology. In WINNER these measurements will be based on the UL and DL synchronization pilots and should be performed by terminals (UTs), base stations (BSs) and the relay nodes (RNs), on the WINNER RAN, but also on legacy RANs, when necessary. Three different types of measurements should be available intra-frequency, inter-frequency and inter-system, the last one should be performed by the WINNER multi-system terminals.
- **Transmitted power**. This is just a report of the transmitted power setting in a precise instant. Pathloss measurements can also be measured as the difference between the transmitted power and the received signal strength. For WINNER RAN should be performed by UT, BS and RN
- **Quality measurements**. This must allow concluding on the quality offer and perceived by the UT and GW and to compare it with the required quality. So it is necessary to do some measurement on user data flow in order to determine QoS level and compare it with thresholds. QoS indicators could be: BLER (block error rate), retransmitted block rate or bit rate at different layers level (for example PHY layer with instantaneous bit rate, MAC layer with throughput or IP layer level), for WINNER RAN should be performed by UT, GW.
- **Cell load**. The cell load corresponds to the currently used resources in comparison with the available by the RAN, at different levels. This shall provide information on the actual cell load, cell load can be measured at different levels radio transmitted and power (PHY layer) or it can be derived from bit rate, number of used chunks compared (MAC user plane), etc. For legacy RANs the cell load definition was presented in [WIN1D4.3] section 3.3.3, for WINNER RAN the input for cell load are presented in [WIN1D4.3] section 6.2, and should be performed by the GW.
- **Terminal velocity and terminal location**. As minimum requirement the system should know to which BS the UT will be attached and to know the coverage area of the serving BS, a more detailed location determination should be performed by a specific entity devoted to UT localization determination**,** with connection to the BSs**,** using timing and received signal strength measurements from different BSs and/or measurements from satellite systems (GPS, Galileo)

The split of measurements between the BS/RN, UT and WINNER RAN could be the following:

- BS/RN measurements:
    - Received Signal Strength (UL)
    - SINR (UL)
    - Synchronization and time measurements (UL).- Observed time difference
    - Total Transmitted Power (DL).- Total TX power compared with the total available
    - Transmitted Power per chunk (DL).- Report
    - Traffic quality measurements (DL).- BLER (but BLER can be derived from SINR)
    - Cell load (at BS level).- Percentage of occupied chunks over total

- UT measurements:
    - Intramode – intra-frequency and inter-frequency (DL).- RSS, SINR
    - Intermode – intra-frequency and inter-frequency (DL).- RSS, SINR
    - Inter-RAT (DL).- RSS, SINR
    - Rx-Tx time difference (DL/UL).
    - Quality measurements (DL): BLER, PER (transport channel)
    - Synchronization and timing measurements (DL): time difference of arrival (TDOA), GPS or Galileo timing measurements
    - Transmitted power per chunk (UL).- Report
    - Traffic measurements (UL, internal measurement): total, average, variance buffer occupancy..
    - Positioning measurements (UL): Cell identification.

- WINNER RAN:
    - SLC and RS buffer load
    - Velocity, direction
    - Handover statistics
    - Network load

The WINNER multi-mode and multi-system terminal should have the possibility to measure the received signal strength of base stations/access points of legacy systems should be taken into account when defining WINNER system. Moreover, the RRM server and the BSs should have the possibility to know what possible cells of other modes and legacy system could handover to have a fast intermode and intersystem handover, when necessary.

4.1.5.4 UT context transfer in active mode

In presence of mobility, the procedure of handover during ongoing service may require a restart of the ongoing service, if the necessary context information is not properly transferred to the new point of access.

In a handover scenario, user and session contexts are needed for advance handover decision making. Context transfer is done in order to make a seamless handover process, i.e. without disturbing the ongoing session and also to make the overall process to be faster, i.e. no need to re-establish and re-authenticate connection. Context information is also important to improve the handover decision making.

However, context transfer procedures introduce additional overheads to handovers possibly affecting the quality of service perceived by the UT and making handovers very critical. Therefore, context transfer must be designed carefully based on certain criteria such as type of ongoing service, or where in the logical entities it happens. Context transfer will be closely related to the policy-based mobility management scheme as described in the previous section.

Each UT that is served by a BS is represented by a user context that is kept in that BS. Each flow that is transmitted via the BS is represented by a flow context in the BS. In the regular case, a UT handover involves the transfer of the user context and the transfer of all UT flow contexts to the target BS by using the BS to BS interface.

The UT comprises all functionality necessary to communicate directly with the network, i.e. a BS or a RN. It contains functions to handle UT mobility in ACTIVE and IDLE states, as well as functionality to perform an initial access to the network. It furthermore contains functionality to initiate a flow establishment.

The User Plane is composed of protocols that implement the actual data transfer services for users, i.e. carry user-data through the radio interface.

User data between GW and the BS is transported by a WINNER specific tunnel protocol. The main purpose of this protocol is to carry the address and flow ID information which has been removed by the header compression entities in the UT and GW, and also to route the flow to the correct BS and RLC instance. This information is used by the BS in the downlink direction to determine the MAC address of the UT and the flow specific RLC instance to transport the data.

In the uplink direction, the same information is used to feed the flow into the specific header compression instance at the terminal side and further to a decompression instance at the network side to restore the IP packet header, and route the flow to the correct GW.

UDP/IP is used as transport protocol for the tunnel. An unreliable transport can be used here because reliable data transmission needs to be guaranteed in an end-to-end fashion. In this case TCP will be used on top of IP between the UT and a communication peer in the Internet. Mobility can be handled in this concept by switching the tunnel endpoints i.e. by sending the UDP packets to the IP address of the target BS.

The WINNER data flow can be described as follows. Each IPCL PDU corresponds to an RLC SDU. The RLC performs segmentation and concatenation of those SDUs and adds an RLC header. The RLC PDUs form MAC SDUs. MAC SDUs from several flows may be multiplexed in MAC. Depending on the amount of scheduled resources, more or less bits are selected for each transport block. The scheduling decision affects the concatenation and segmentation in RLC and the multiplexing in MAC.

One issue that needs to be taken into account, when UT changes its state from idle to active mode, is how the UT is assigned an IP address in the active mode while it has only identification in idle mode.

Context transfer from source BS to target BS in active mode can be carried out in two ways, i.e.:

- Through air interface: There is a channel which is assigned for the transferring the context;
- Through the core network: in this way, there are two transport protocols involved i.e. SCTP is used in the CP and UDP in the UP.

The user context contains the user preferences. Support of the user preferences will depend on the capabilities of the UT and the type of service. If the capabilities change, the degree of support of the user preferences may change too. Table 4.2 shows an extended content of Table 4.1 which contain not only static user context, but also including the dynamic ones.

Module Name	Contents	Type	Kept in
User	User ID General Preferences Context by default	Static	GW_{CP}
Context	Context ID Context Restrictions User preferences	Static	GW_{CP}
Service	Service ID Suspension status Subscribed QoS	Static	GW_{CP}
Subscriber	Subscriber ID	Static	GW_{CP}
Service provider	Service provider ID	Static	GW_{CP}
Service profile	Service restrictions User availability User settings	Static	GW_{CP}
Terminal	Terminal category Hardware configuration Supported service list	Static	GW_{CP}
Access network	Network ID Static network configuration	Static	GW_{CP}
Access session	Connectivity status Network PoA/location	Dynamic	GW_{UP}

	Available QoA		
Service session	Service status Statistics/accounting data	Dynamic	BS
User's RLC Flow Context	RLC SDU RLC PDU (advance function)	Dynamic	BS

Table 4-2: Types of Static and Dynamic User Context

In the following the user context transfer is described for the scenario of Mixture HO.

There are two phases of data transmission that should be continued transmitted in the mixture HO (Figure 4-20):
- During radio handover, a tunnelling protocol, e.g. GTP at the user plane, is used to forward the retransmission packets to the target BS;
- When target BS registers at the attached target GW_{UP}, the new arrived traffic of the UT will be directly forwarded to the target BS;
- The source GW_{UP} forwards the new packet to the target GW_{UP} also through the tunnelling protocol, and at the same time target BS forwards the "router advertisement" of the target GW_{UP} to the UT;
- After forming the new CoA (Care of Address), the UT sends the binding updates to the CN (corresponding node). The term CN means an endpoint node where the UT is connected to. For example, when a UT is accessing the web, the CN of this UT is the web server it connects to. When the UT establishes a VoIP session then the CN is another UT which is in the same VoIP session;
- After the CN (corresponding node) has been updated with the new route, the new packets will be directly routed to the target GW_{UP}.

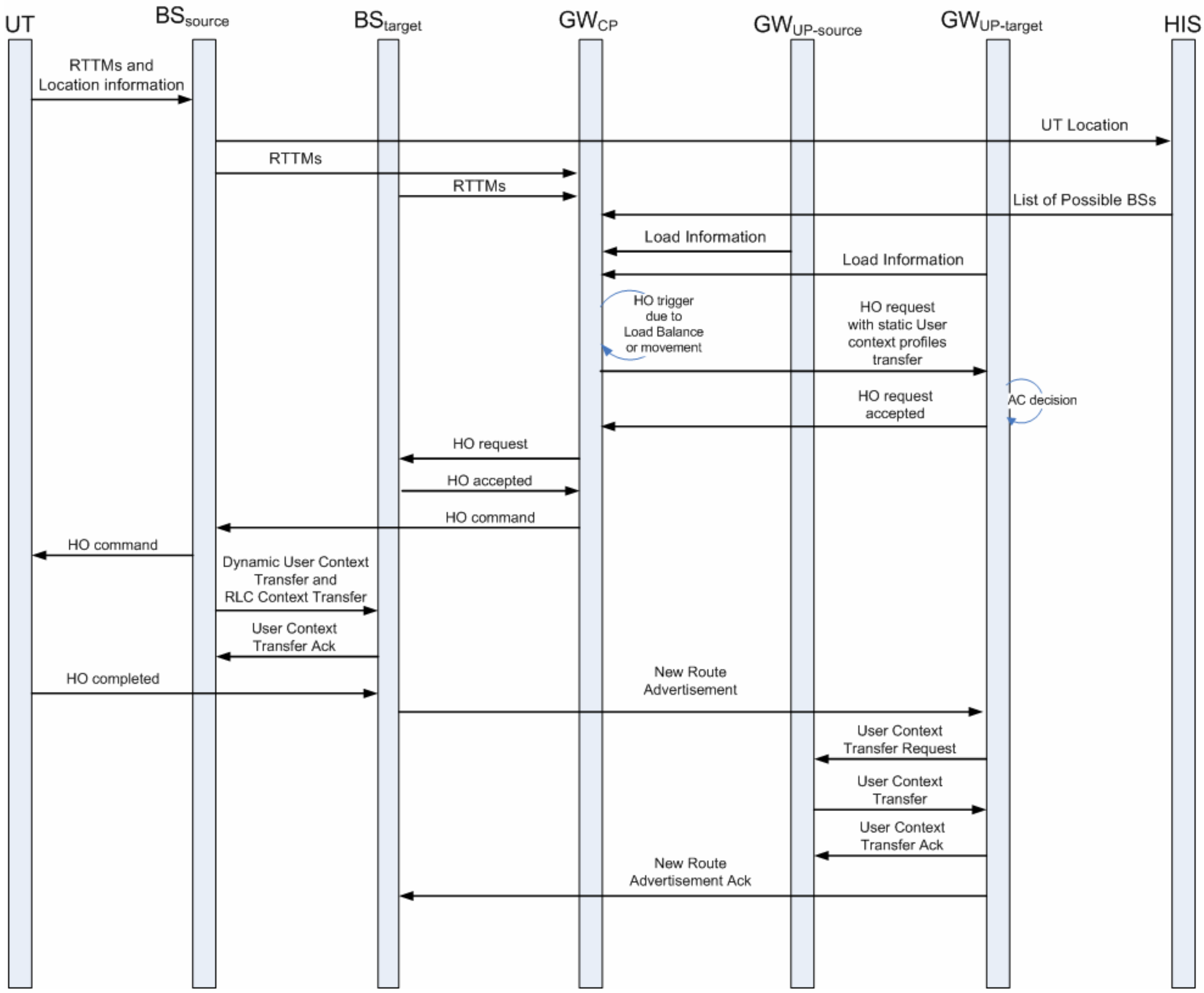

Figure 4-20: Signalling flow and context transfer in Mixture HO

4.1.5.5 RLC SDU context transfer during handover

RLC-ARQ (outer ARQ) terminates in the BS and in the UT, which implies that RLC-ARQ cannot recover data loss due to BS-BS handover. Reliability is important for higher layers, e.g., for applications that use TCP, such as web surfing and e-mail. Data loss due to handover should be avoided, since it reduces TCP performance significantly. For high bandwidths, the performance degradation is even more severe [Bajzik07]. Increasing error rates and round trip times also reduce performance. The reason is that bandwidth, error rate and round trip time have an impact on the pipe capacity. If the pipe capacity increases, then more RLC data is buffered in the source BS and, hence, more data is lost if buffers are discarded during handover [Sachs06]. A solution to avoid data loss due to handover is to forward data from the source BS to the target BS during handover. Forwarding of buffered RLC SDUs, RLC SDU context transfer, is recommended as a required feature in WINNER. The following scheme for RLC SDU context transfer is proposed to avoid data loss due to handover:

In DL
1. Before handover, the RLC receiver in the UT transmits a status message to the source BS.
2. During handover, the RLC sender in the source BS forwards all buffered RLC SDUs to the target BS.
3. After handover, the target BS transmits all RLC SDUs that were forwarded from the source BS to the UT.

In UL
1. Before handover, the RLC receiver in the source BS forwards all successfully received RLC SDUs to IPCL in the GW.
2. After handover, the UT transmits RLC SDUs.

In order to avoid inconsistent states between sender and receiver before handover, we propose that, in DL, the RLC receiver in the UT is instructed from higher layers to transmit a status message to the source BS before handover. If the status message is successfully received, some RLC SDUs may get acknowledged and less RLC SDUs may need to be forwarded. If the status message is lost, RLC SDUs that have already been successfully received may unneccessarily be forwarded and retransmitted.

Forwarding of RLC SDUs was chosen for 3GPP LTE [3GPP300]. Simulations in [Racz07] indicate that TCP performance is improved with RLC SDU context transfer. For WINNER, we propose RLC SDU context transfer as a required function, and RLC PDU context transfer, in addition to RLC SDU context transfer, as a value added function. RLC PDU context transfer is described in Section 4.2.6.

4.1.6 Inter-system handover

4.1.6.1 Handover process

It is assumed that when a UT camped in WINNER could handover to a legacy RAN when it lost the coverage of WINNER system, this situation relevant especially in the initial deployment with very limited number of WINNER BS, congestion in WINNER cells. The work related with intersystem mobility management and especially intersystem handovers was presented in most of the previous deliverables in Phase I and Phase II. This work was updated in the WWI Cross Issues System Interfaces workgroup and here we will present the output of this work into a global heterogeneous RRM mobility management function.

Mobility management in the WWI concept is related to mobility management between different access networks, since all the projects are working on enabling cooperation of networks, but each one from a different perspective, i.e. WINNER is working towards making the new network able to cooperate with the legacy networks. Some of the common requirements from all the projects for mobility management are:

- **Handover:** Fast and seamless handover execution and need for seamless reconfiguration for the execution of the handover.
- **Triggers:** Either for handover and/or general mobility actions, triggers have been acknowledged in all the projects. The actual triggers largely vary and can include physical layer-related triggers, context information, user preferences and profiles, location information more implicit cases such as adaptation/monitoring of service to system changes and intelligent triggers.
- **Context and location awareness:** The availability of context and location information to the mobility management processes is another requirement that, if utilized, expedites and improves important performance criteria
- **Planning:** This relates to the actions that are performed in response to determination that a handover is imminent or in an execution stage. The requirement also demands a fallback to unplanned handovers in case a planned handover fails.

- **Coping with legacy systems and multiple operator deployments:** Support for different technologies for vertical handover, could relate to seamless execution of reconfigurations for terminals attempting to matching to the spectrum and network standard. One extreme case is the network reconfiguration due to mass handover requests from terminals with the same radio mode. For example, terminals should be able to become reconfigured when entering into a new access network that belongs to the same or even a different operator. The need for reconfiguration arises from the fact that different networks, especially those belonging to different operators, may use a different set of mobility management, RRM and QoS protocols. Thus, the user equipment needs to be able to be adjustable to the new environment in a seamless way. Obviously, this requires the ability to discover early any incompatibilities between the current configuration and the protocols used by the new network, locate and download the appropriate set of protocols, and reconfigure the protocol stack appropriately.
- **Mobility as diverse architectural component:** This includes all of its aspects: session/application, traditional, resource expenditure, load balancing, multi-homing, security, all variants of mobility: endpoints, sessions, flows, interfaces, network/groups, flexible spectrum utilization, group and signaling managements..

A WWI functional architecture for the handover process part of the mobility management is shown in Figure 4-21. The different functional entities of all the projects related to inter-RAT handover are presented and also the connections between these entities. This figure is composed of several layers, starting (from bottom to top) from the physical layer and ending at the application and service layers. The entities included in the figure have different colors to represent the project they come from. The blue boxes are the entities from Ambient Networks, the green are from WINNER, the yellow from E^2R, the orange from MobiLife and the brown from SPICE. This picture does not show the overall WWI architecture for the mobility management topic, but it is a good example of the inter-RAT handover process and how the entities of the projects could cooperate in a unified way of handling an Inter-RAT handover. The central entity without color is the global RRM entity (named HRRM) that is responsible for the RRM decisions, the RAT selection, the load balancing etc. This figure also shows many triggers for an inter-RAT handover, either from the network layer (from SRRML or SRRMW) or from the service layer from service or context information. The newest addition is a vertical cross-layer entity related to fuzzy logic, which is described in the next section.

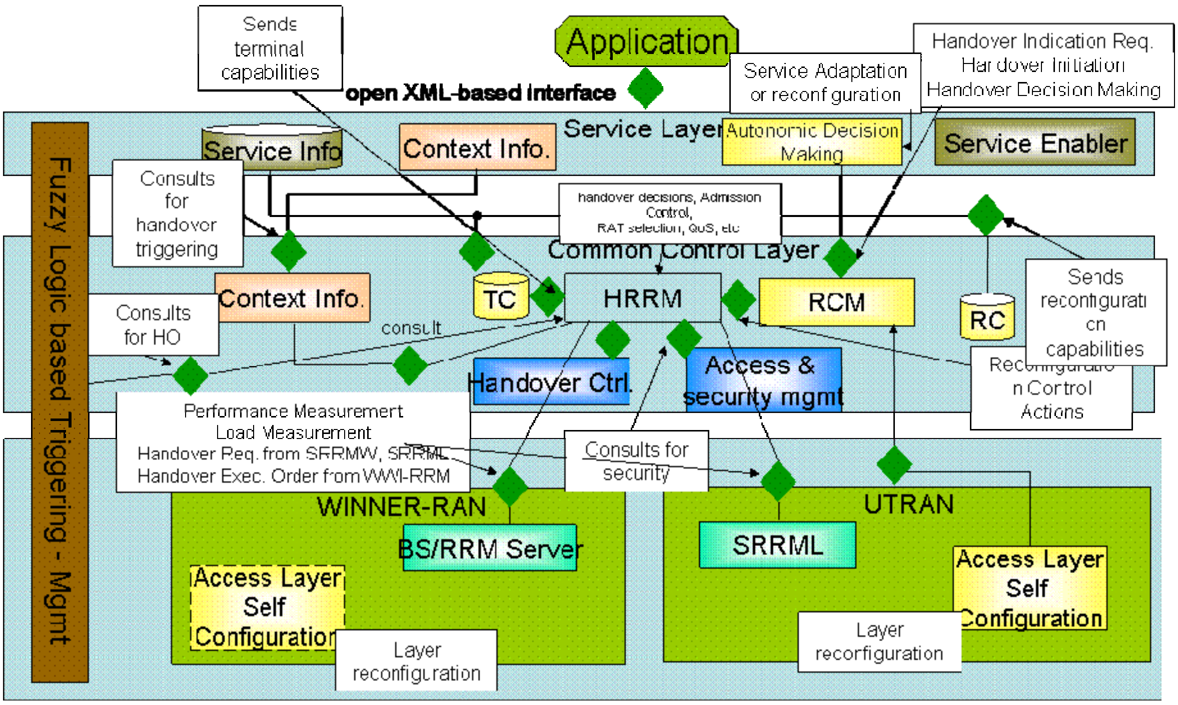

Figure 4-21: Example of Functional Architecture for Inter-RAT handover process

In Figure 4-22 is presented a signalling diagram for the WINNER RAN and reconfigurable RAN HO request to HRRM for Inter System HO. This is a handover of an active session from a WINNER RAN to a second RAN (which is not a WINNER RAN and might either be reconfigurable or not). As this is a typical inter-system scenario the functions in the common core network are mostly concerned about the realization of this procedure.

The handover scheme employed by the WWI network model relies on the assistance of the mobile terminal to provide information about the access quality (e.g. C/I ratio) to the access selection function in the core network. Based on this information and the core network's knowledge about the end-to-end path, the access selection function in the core network eventually initiates a handover request. The first step in the execution of the handover is the selection of the appropriate handover tool, as we operate in a mult-technology environment and different access networks are typically employing different solution and protocols to execute a handover. The following actual handover execution follows a make before break scheme controlled by the handover execution function of the core network.

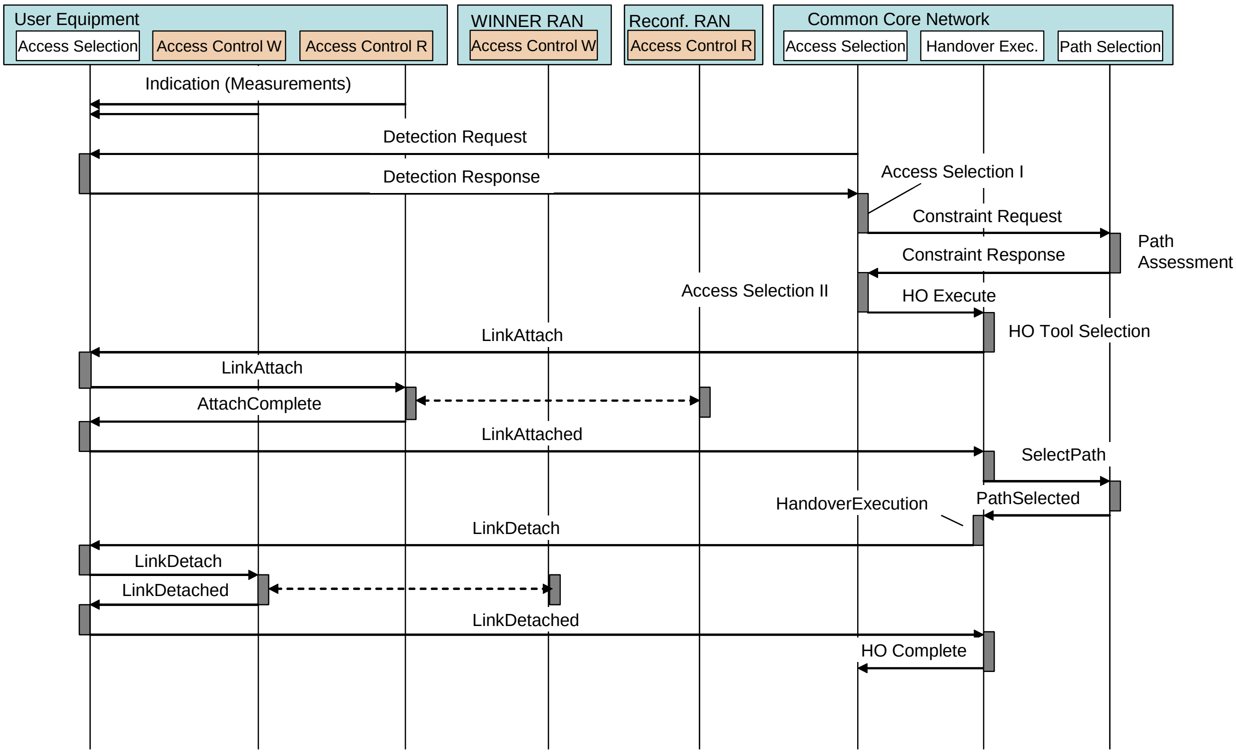

Figure 4-22: intersystem HO signalling

4.1.6.2 UT measurement reporting

To choose the most suitable radio access technology, a fundamental aspect is to assess the quality that different RATs will provide to the users. This quality indication is basically obtained from measurements on the current and target networks. RRM mechanisms in B3G systems will be able to use more metrics and information as inputs than the current RRM algorithms, since information will be exchanged between networks and new metrics are becoming available.

Current handover mechanisms are only based on received power either in cellular systems or in WLAN networks. Congestion control algorithms in UMTS use mainly power measurements and service attributes to control access and load. On the contrary, in B3G systems various metrics coming from all the networks available can be used as inputs of any type of RRM algorithms.

Some of these new metrics or information that is foreseen as useful to improve the RRM performance are listed below:

- Cell load, free capacity
- Location
- Velocity
- User's environment (indoor, outdoor, etc.)
- Terminal capabilities
- Handover statistics

Measurements can be performed either in the base stations or in the user terminal. Usually RRM decisions are taken by the network, therefore mechanisms must exist to report measurements from base station or terminal to the network. The reporting may be periodic, event-triggered or on demand.

4.1.7 Flow admission control

Admission Control (AC), which is the mechanism that receives the requests for new flows (whether they come from a new user or from ongoing users) and checks if the users are authenticated to the network and if the network has sufficient resources based on the requested resources by the new session.

There are several criteria that the Admission Control algorithms use for accepting or rejecting a flow. Many algorithms are power-based, which means that they use periodic measurements of the transmitted power, computing the interference at the user's receiver and based on that it makes the decision of admitting or rejecting the user. In throughput-based algorithms, the throughput that can be delivered by the system is determined according to some dimensioning calculations, assuming some conditions in the system. There are also algorithms that use the equivalent capacity of aggregated traffic, which is an estimation of the arrival rate of a class of traffic. Other algorithms check the load of the system not to exceed a pre-defined threshold and the bandwidth and delay constraints of each flow, according to the current system's data.

In mobile environments it is not adequate to admit a new call only based on the status of the current cell, where the call is being generated, because when the user attempts to move from the original cell to a next cell, there may not be sufficient resources in the destination cell for accepting the handoff. This may result in dropping the call and increasing the call dropping rate. This results that an efficient Admission Control algorithm should check the available resources in the adjacent cells, except in the current cell.

Moreover, not all sessions have the same characteristics and requirements. For example a voice session (telephony) has very low delay and bandwidth requirements, though e-mail delivery is very tolerant to delay and it needs more bandwidth available. That's why in radio access networks there are defined different service classes for the users according to the services the networks will offer and to the different characteristics of each service. Future wireless networks will have to consider many service classes to meet future user requirements for best quality of service offers, because they want to offer the best quality of service to the users and they have to meet the future users' requirements. The service classes considered have different delay, throughput and BER requirements. The different service classes of the RANs will be acknowledged by the AC algorithm, in terms of resource allocation and especially in terms of prioritization. Different service classes will have different priorities in the algorithm. The criteria for each class's priority should be based on the characteristics and requirements of each class, i.e. the delay sensitivity, the bandwidth requirements, etc. A class with high priority should be checked (admitted or rejected) first than a low priority class, although the low priority class arrives first, or a high priority class could be admitted though a low priority class will be rejected. Due to the nature of the future service classes there could be services that should have higher priority irrespectively to if the call is new or from handover. For example if in a network is defined a service class for "emergency calls" (i.e. calls for police or ambulances during a car accident), these calls (which are new calls) should have the highest priority even from handover calls.

In Figure 4-23 the actions of a flow admission control defined for the WINNER system are presented.

The different options of the WINNER RRM architecture (described in the next chapter) are also affecting the admission control procedure, so there are three different options, namely centralized, distributed and scalable/hybrid admission control.

Centralised AC can only be done when a central entity (like the RRM server in WINNER) exists in the network. The RRM server receives periodically RTTMs (Real Time Traffic Measurements) from the BSs and the GWs, so it has permanently the knowledge about the load of the BSs and the GWs. When a user tries to change his state from idle to active, the AC is handled by the RRM server, which, based on the load information, can find the most suitable (and not overloaded) mode and BS to serve the user.

Distributed AC means that the AC decisions are only taken by the BS that the user is trying to connect. This is very similar to the existing AC procedure for independently operated systems. If the user is not admitted to that BS he will try to be admitted to other BS in the same area. This approach is very fast and does not need a lot of signalling exchange between the network entities.

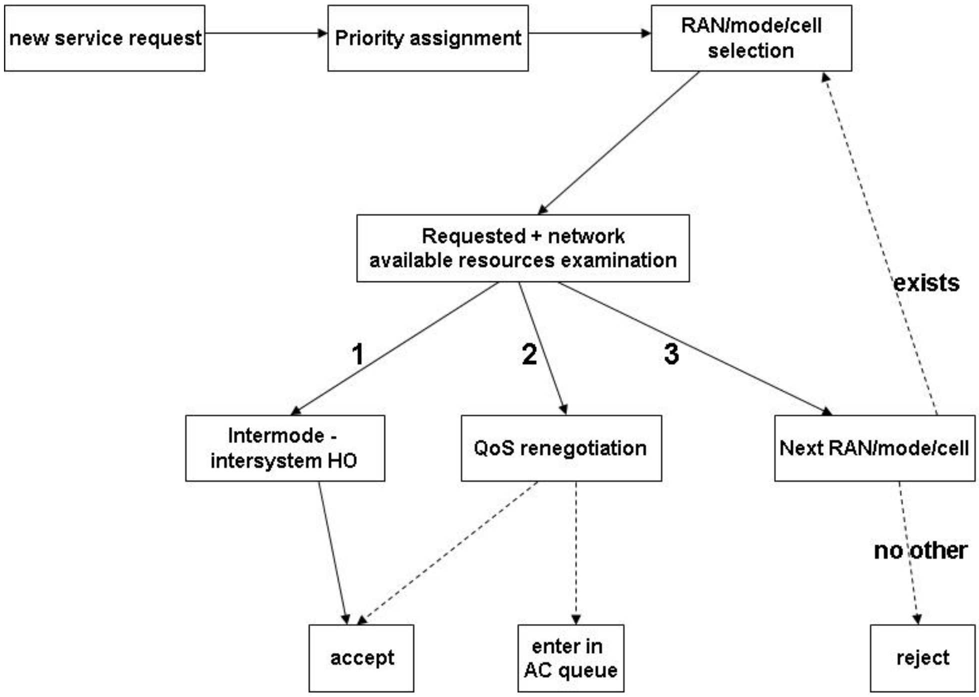

Figure 4-23: flow admission control

In order to exploit the advantages of the two previous options, in WINNER we have adapted the approach of scalable/hybrid admission control. In this approach the AC is being handled by the BS, but it can request assistance from the RRM server to take a better decision (see Figure 4-24).

The hybrid/scalable AC solution is usually fast since the BS takes the decision and does not need so much signalling between the entities, since the information exchanges between the BSs and the RRM server could be performed only on-demand and not all the time.

The current BS is the one that communicates with the neighbour BS in order to take the decision about admitting the new user request. The current BS also communicates with the RRM Server in order to get advice for the target BS, so as it can take the best decision for the user. The BS can also complete the AC procedure without asking the RRM Server for advice and this can happen in low load situations. The signalling process for the scalable/hybrid Admission Control is presented in Figure 4-24 . Same as in Figure 4-23, the dashed lines mean that the signalling is optional, when the RRM Server is to be taken into account.

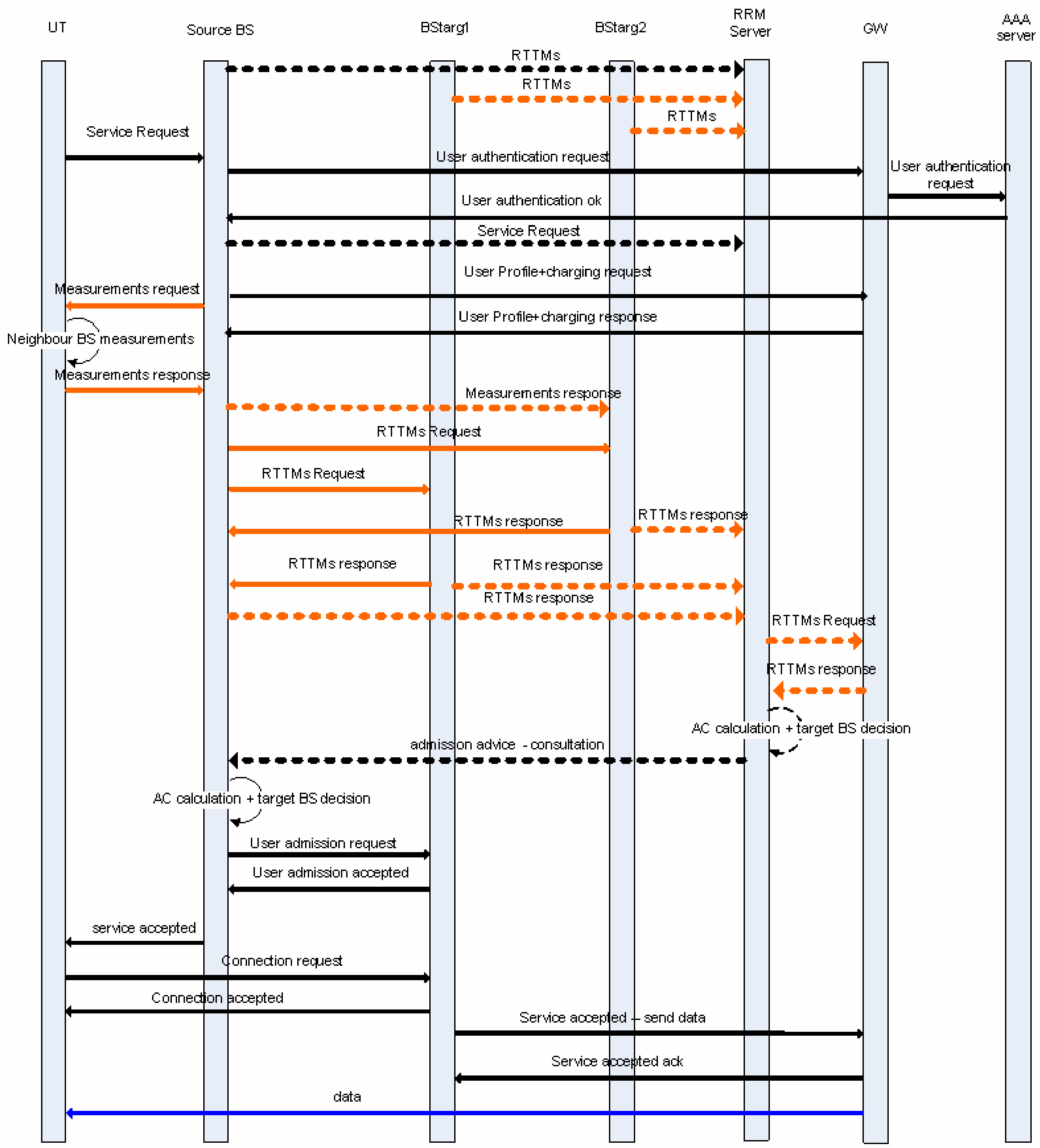

Figure 4-24: Hybrid / scalable Admission Control signalling process

4.1.8 Flow establishment maintenance, reconfiguration and release of flow classes

After one data flow is admitted by the admission control (AC) process, the RRC layer and its messages are in charge of the control signalling for flow establishment, maintenance, reconfiguration and release. The RRC layer may receive a flow setup/reconfiguration request from the peer RRC layer, a higher layer or the IPCL layer. The procedure of flow setup defines the configurations of the radio protocol layers (Figure 2-1and Figure 3-1) that are necessary to provide a class of service that the flow requires. For the setup of a flow, first the IPCL layer has to be configured, for example, whether the header compression scheme shall be enabled and which header compression scheme shall be selected for the flow in question. Also, whether or not to enable the IPCL feedback information has to be decided. Finally the process of transfer of user data between IPCL entities (see section 3.3) should be started, this service utilizes RLC layer to transfer user data.

After the configuration of the IPCL layer, the RLC layer has to be configured. A logical channel is interposed between the RLC layer and the MAC layer (see Figure 3-1). One data flow uses one independent RLC entity which then uses one logical channel (see Figure 3-1), On the RLC level, first its operating mode has to be selected, be it the acknowledge mode (AM), unacknowledged mode (UM) or the transparent mode (TM). Also the RLC has to decide which logical channel to be used and hence the interface to the MAC could be determined.

The transport channels are interposed between the MAC and the PHY layers. A variety of transport channels are defined for different purposes such as broadcasting, paging, forward access or transmitting normal unicast data. Multiple logical channels could be multiplexed into a single transport channel for transmission. Likewise, multiple transport channels could be multiplexed into a single physical channel and demultiplexed on the receiver side. The transport channel and physical channel parameters are included in the flow setup/reconfiguration procedures but can also be configured separately with transport channel and physical channel dedicated procedures.

After the flow transmission is terminated or the radio handover is completed, the RRC would release this flow and reconfigure the related lower layers and their interfaces. Allocated radio resources would also be released.

4.1.9 Flow control

There are two interfaces that needs flow control, one is the GW-BS interface, the other is the BS-RN interface in case the UT is connected via the RN. On both interfaces the flow control mainly refers to the downlink.

4.1.9.1 GW-BS flow control

Due to the capacity difference between the GW and BS, overflow of user-plane data at the BS might happen for a single or a group of UTs. If the BS detects that the buffered data of a particular flow is rapidly increasing and approaching the buffer limit, it could issue an explicit signal to the GW to suspend the forwarding to the BS, in addition to preventive buffer management policies such as RED (random early dropping). This explicit signaling conveys two fields (see Figure 4-25), one is necessarily the identifier of the IPCL data flow (or of the UT) that is approaching the buffering limit, and the other is a command that requests the GW to suspend or to resume the data forwarding. Such command could be identified by the IPCL layer on the GW, so it would temporarily hold its PDUs in buffer, instead of forwarding them immediately to the RLC on the BS after the processing of SDUs. If later the congestion state for the flow at the BS is alleviated, the BS could again send a recovery message that informs the GW to increase the forwarding rate.

Instead of having a simple flag to stop/resume the GW-BS data forwarding, the BS could also advertise a receiving window size to the GW, so that the IPCL layer on the GW would only forward as much data as allowed by the receiving window of the BS. Such advertisement of receiving window by the BS to the GW could be updated periodically or event driven, e.g. in case of handover or other situations that suddenly decrease the air interface data rate for a particular UT.

During handover process, the BS-GW signaling could also be utilized to help maintain the in-sequence delivery of IPCL PDUs from the GW to the BS. As described in Section 4.1.5.5 and Section 3.6, context transfer has to be executed between the source BS and the target BS. The buffered RLC SDUs and preferably also the RLC PDUs shall be forwarded from the source BS to the target BS. The source BS also has to notify the GW about the switching of the data forwarding path after the handover to the target BS is completed. Before the notification of the path switching, the GW would still forward the IPCL PDUs to the source BS, which then needs to be tunneled by the source BS to the target BS. If the forwarding of those data is not finished before the path switching, new IPCL PDUs from the GW might arrive at the target BS before some of the previous IPCL PDUs from the source BS. The RLC layer on the target BS can not recover the original order of the IPCL PDUs, unless it could probe into the IPCL header to obtain the sequence number information, which violates the layering paradigm. A possible solution is to utilize the aforementioned GW-BS flow control signaling. The source BS could request the GW to stop forwarding further IPCL data to it, to avoid unnecessary tunneling of the PDUs to the target BS. In addition, the source BS should notify the GW about the path switching after it has forwarded all the buffered RLC SDUs/PDUs to the target BS. As a result, the forwarded data from the source BS always comes earlier than further IPCL data from the GW to the target BS, so that in-sequence delivery could be preserved. The whole handover process with downlink GW-BS flow control is shown in Figure 4-25, where the XON/XOFF represents the command to suspend/resume data forwarding from the GW to the BS.

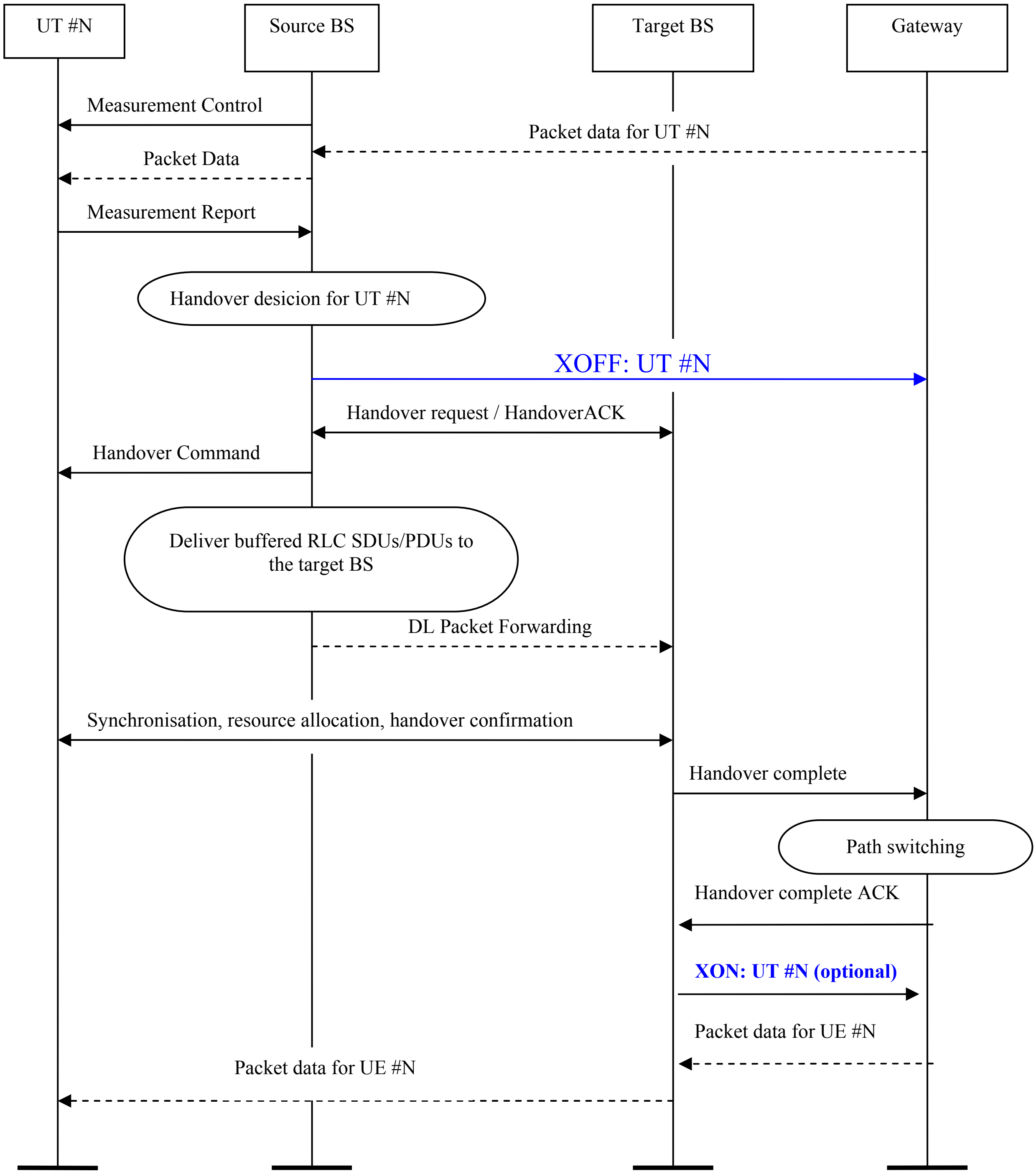

Figure 4-25: GW-BS flow control signaling during the handover process

4.1.9.2 BS-RN flow control

With fixed RN and the line-of-sight connection between the BS and the RN, the data transmission between BS and RN should be stable and of high data rate. In contrast, due to the fading condition, the RN-UT link is unstable and usually of lower data rate than the BS-RN link. Although the downlink transmission is controlled by the scheduling and resource allocation on the BS, without specific flow control, overflow or underflow situation at the RN buffer might happen. The principle of BS-RN flow control is to keep an appropriate level of buffering status on the RN. The flow control mechanism should operate on a per-flow basis, which controls the BS sending rate for the RN with the awareness of the RN buffer status.

It is suggested that the RLC layer should also be implemented on the RN, so that two separate RLC connections would be in place for the BS-RN and RN-UT link (see Figure 4-26). The two RLC connections could conduct independent ARQ (if RLC AM mode is configured) on each individual radio link, which makes the error recovery more efficient than the case of single RLC over the BS-RN-UT path. Nevertheless, the two RLC connections could share the same sequence number of the RLC PDUs, so that RLC SDUs don't have to be restored on the RN and again segmented for the RN-UT link. As shown in the following figure, for the RLC AM mode, with shared sequence number of the two RLC connections, the RN could forward the status report from the UT back to the BS, in addition to its own RLC receiver status report. With the awareness of both the RN status and the UT status at the RLC layer of the BS, the buffering status on the RN could be implicitly calculated by the BS. The BS could then adapt its sending rate for the RN according to the calculated RN buffering status, by adjusting the scheduling/resource allocation at the MAC layer. The pace of sending both status reports to the BS could be controlled by the RN, since it could poll the UT to obtain UT status whenever needed.

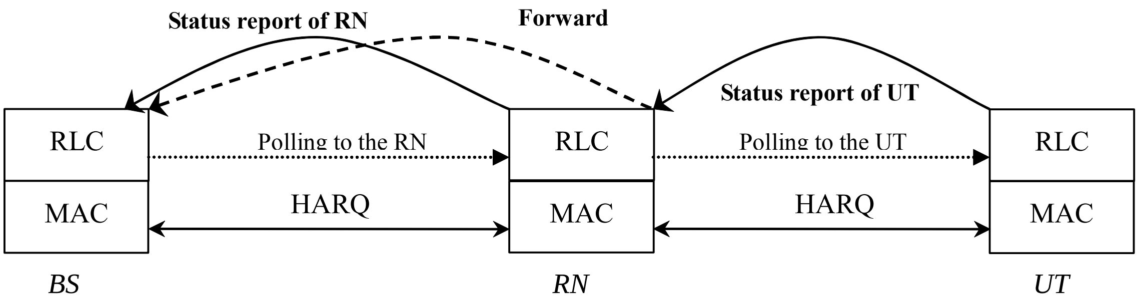

Figure 4-26: Status report forwarding by RN to the BS for the downlink transmission

In case of transmission by RLC UM mode, no status report is defined and hence the buffering status of RN can't be known at the RLC layer of the BS. In this case, a specific flow control signaling might have to be defined at other layers. It could be a special RRC message, which asks the BS to increase or decrease its sending rate to the RN for a particular logical channel. Either it could be a special MAC layer message, which requests the MAC scheduler at the BS to adjust the scheduled data rate for the particular logical channel. Nevertheless, we predict that the RLC AM mode would be the main operating mode, due to its ARQ functionality and the efficiency enabled by ARQ/HARQ interworking (Section 5.5.1)[4].

4.1.10 Load/ Congestion Control

Admission Control algorithms are designed to make decisions about the new session requests, in order to mantain the load of the networks under a certain threshold. Though, if for different reasons, the load of the network exceeds that threshold, then the network experiences an overload/congestion situation. Network congestion control is a very critical issue and has high priority, especially given the growing size, demand, and speed (bandwidth) of the increasingly integrated services networks. Designing effective congestion control strategies for future wireless networks is known to be difficult because of the complexity of the structure of the networks, nature of the services supported, and the variety of the dynamic parameters involved. Congestion situations can be caused by saturation of network resources such as communication links, channels, throughput, etc. For example, if a communication link delivers packets to a queue at a higher rate than the service rate of the queue, then the size of the queue will grow. If the queue space is finite then in addition to the delay experienced by the packets until service, losses will also occur. Networks need to serve all users requests, which may be unpredictable and bursty in their behaviour (starting time, bit rate, and duration). However, network resources are limited, and must be managed for sharing among the competing users. Congestion will occur, if the resources are not managed effectively. The optimal control of networks of queues is a well-known, much studied, and notoriously difficult problem, even for the simplest of cases.

The basic result of a congested network is the degradation of the network performance. The users are experiencing long delays in the delivery of the packets, perhaps with heavy losses caused by buffer overflows. The jitter and delays values are then very high and the network available throughput can be close to zero. The degradation of the network performance, the delays and the packet losses result to retransmissions of the lost packets, which in turn results to a waste of the available throughput, which is consumed to retransmit the lost packets.

There are many ways to detect or sense a congestion situation in the network:

[4] The given flow control protocol is considered as one realization of the inter BS-RN flow control. Alternative approaches of flow control under this context can be found at [WIND6114].

- Packet loss sensed by the queue as an overflow, by destination (through sequence numbers) and acknowledged to a user or by sender due to a lack of acknowledgment (timeout mechanism) to indicate loss.
- Packet delay, which can be inferred by the queue size, observed by the destination and acknowledged to a user (e.g. using time stamps in the packet headers), or observed by the sender, for example by a packet probe to measure Round Trip Time (RTT).
- Loss of throughput, which can be observed by the sender queue size (waiting time in queue).
- Other events, like increased network queue length and its growth or queue inflow and its effect on future queue behaviour.

In Figure 4-27 an algorithm for Load/congestion control in WINNER is being presented. This algorithm is split in three phases, the congestion detection phase the congestion resolution phase and the congestion recovery phase. In the Congestion Resolution phase five different steps for decongesting the network are presented. These steps are numbered from 1 to 5. These numbers show a kind of prioritization, in terms of which step is applied first by the network. The first action of course would be to reject any non – emergency new requests. Then the network will do some spectrum management techniques in order to gain more resources in the congested area. Then other techniques like handovers, resource renegotiation and dropping high loaded flows are applied. These priorities are assigned to these steps taking into account their effectiveness on the load reduction and on the user's perception/

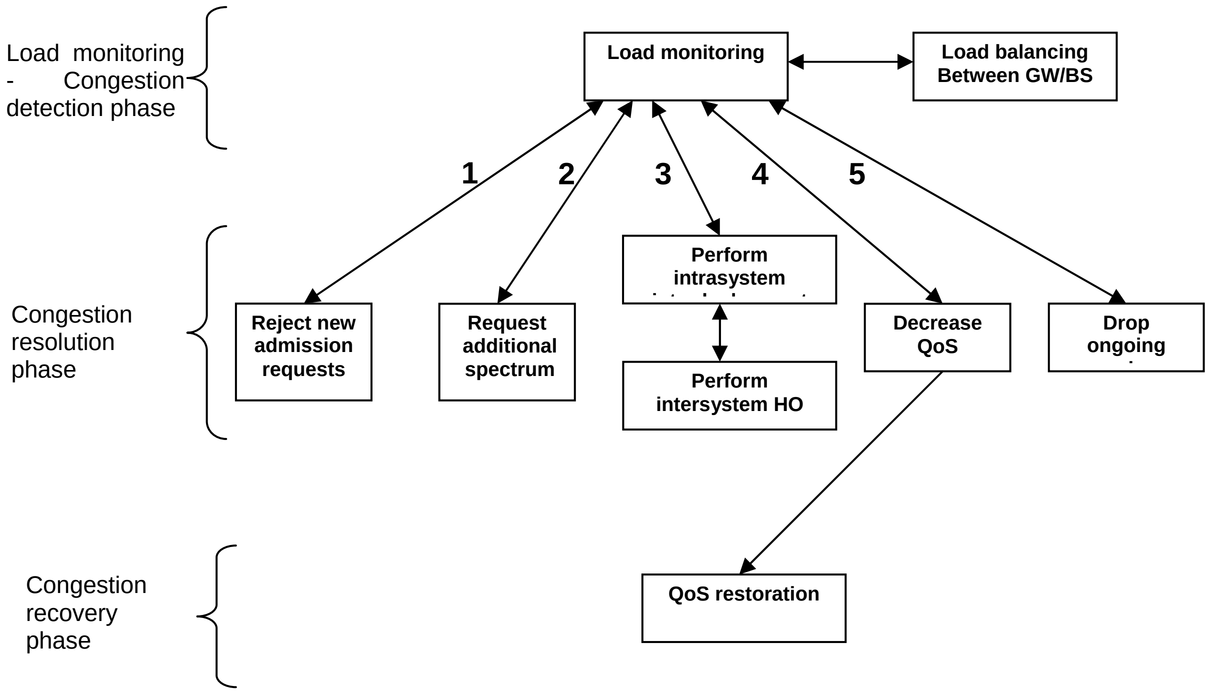

Figure 4-27: Load Control functions interactions

In Figure 4-27 the interworking between Load and Congestion control is presented, according to the value of the load in the network. In each of the states of the load there are also defined the functions that take place and the priority in these functions.

WINNER supports radio context dependent optimal function selection, which is supported by the proper inter-funciton cooperation. The classification, differences and interworking among WINNER RRC/RRM functions are further explained in Section 5.7.

4.1.11 RRC security functions

The overall security scheme messages is composed by two layers, as shown in Figure 4-28, the first one protecting the radio transmission in the RAN Access Stratum (AS), specifically the RRC messages between the BS and the UT, send by radio and the second one protecting the signalling between the UT and the data bases with UT data (HSS-AuC) belonging to NAS.

In section 3.8 was presented the ciphering at user plane (IPCL layer) level, this ciphering is also used by NAS communications, the terminating nodes in this case are the UT and the GW (and also the HSS in the case of NAS signaling). Different keys will be used for ciphering and integrity protection of RRC messages and NAS/user plane messages. These keys should be refreshed each time a user is reauthentified.

The objective of this two layers security (in UMTs only there was a security layer) is to reduce the effects of the compromised radio transmission (first layer) from the second layer, now the RRC layer terminated in the BS (in UMTS it is terminated in the RNC, a distant a more innacesible node). Assuming that radio access will be the preferred mode to attack the network, especially considering that the BS will have more functionalities, rogue BSs can used to extract data from the user simulating operator's network behaviour. The aim is that in case of first layer security would be broken this could have a limited effect on the security of the second layer.

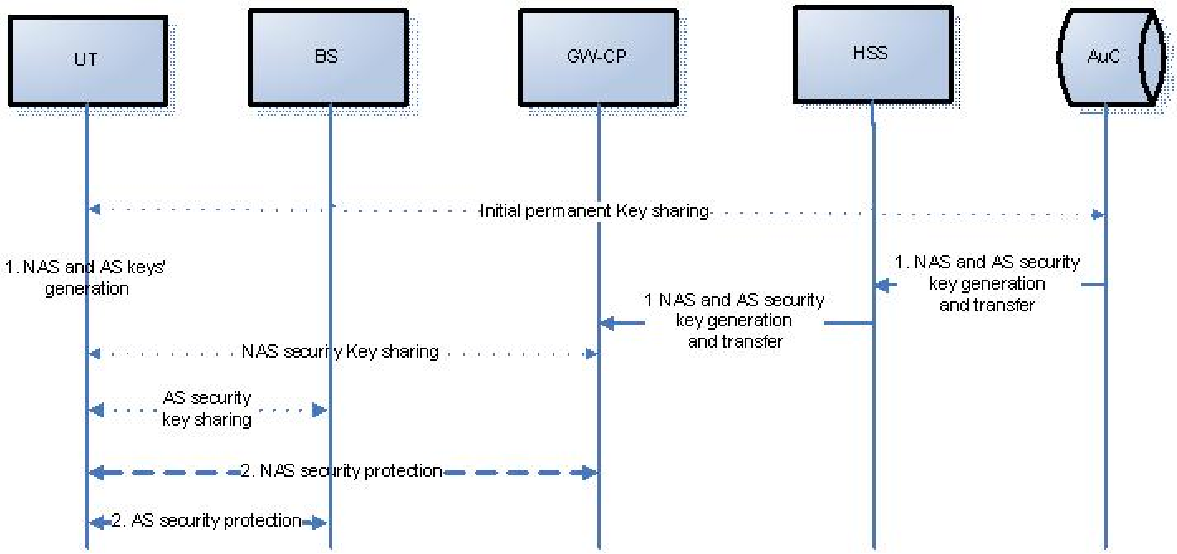

Figure 4-28: RRC layered security scheme

4.1.11.1 Integrity protection of RRC messages

Integrity protection for RRC messages should be performed for preventing the insertion, modification, deletion and replay of RRC signalling messages. Since the keys used for the integrity protection are shared between network/GW-CP and UT, mutual message originator authentication can also be ensured by checking the intregrity of packets.

CP integrity protection is performed between BS and UT for AS signalling and between GW and UT for NAS signalling. NAS signalling can be used for authentication, key agreement, algorithm negotiation between CN and UT, thus it is mandatory to apply integrity protection to them. Moreover, the integrity of most of AS signalling messages should be also mandatory. A MAC (Message Authentication Code) code will be used for the integrity protection of CP signalling, but the algorithms are still TBD (UIA2 and AES shall be used).

4.1.11.2 Ciphering of RRC messages

The ciphering can guarantee the confidentiality of RRC messages. Thanks to it, the following vulnerabilities can be avoided: eavesdropping of messages, unlawful trace of users, disclosure of user identities, unlawful interception of messages, etc. Moreover, ciphering is also able to ensure to some degree the integrity and authentication as it proves the possession of the corresponding keys shared by the communicating entities.

CP ciphering for NAS and AS signalling are between different entities and use different keys. For NAS CP traffic, a ciphering key is shared between GW and UT, whereas for AS CP traffic, another ciphering key is shared between BS and UT. UEA2 and AES algorithms shall be used.

4.2 Advanced functions

Hereafter, are presented some advanced functions that do not belong to the basic system concept, but in some circumstances and scenarios could increase system performance or to present advantages at system level, depending of system implementation and environment.

4.2.1 Hybrid and scalable network

In section 5.1 is presented the principles and implications of the proposed scalable network architecture in detail, in [D4.8.2] were presented the advanteages of centralized RRM that, in some scenarios, could justify the increase of complexity to obtain performance gains and advanced functions.
Basically it is proposed an innovative and hybrid approach, with a baseline of pure distributed RRM architecture, like many current systems on standarization process (3GPP LTE, WiMAx Mobile) where much functionality has been moved to the BS (handover, admission control, load control, etc.), but it have been identified several advanced functions that need a central node to coordinate sets of BSs (inter-cell frequency reuse, inter-cell load balancing, etc) that justify the use of the optional centralised RRM server in situations of medium-high load in the network and also when advanced functionalities or optimal performance are needed. In this situation central coordination will offer superior performance, therefore it is proposed a hybrid and scalable architecture.

4.2.2 Distributed and centralized handover and hybrid handover

In this section the different options for handover related to the WINNER RRM architecture will be presented. In WINNER the baseline is a distributed RRM architecture, but there also they are under consideration a centralised node like the RRM Server to execute RRM mechanisms in a centralised architecture option or having this central node assisting the BS in taking the RRM decisions in a hybrid solution. There are many differences in the handover process for the different cases of distributed, centralised and hybrid architecture. In the following sections the three different approaches will be presented and analysed.

In the distributed handover version, the handover process is being handled exclusively by the current BS that the UT is connected to. There is no central node taking or assisting the decisions. The information and the signalling run through the current BS, which communicates with the neighbour BSs through the IBB interface. The information exchanged between the current BS and the neighbour BSs is related to real time traffic measurements (RTTMs) and mostly related to load information, so the current BS can take the decision and make a list of candidate BSs for the handover of the UT. The target BS can be connected to the same gateway (solid line of UP data) or to another gateway in the same pool (big dashed blue line of UP data). The UT is moving inside the WINNER network and performs a handover between two base stations that are controlled by the same pool of gateways. The current BS takes the HO decision and communicates with the target BS to complete the handover.

In the centralised option of the handover, there is a central node that takes care of the RRM decisions. This node in the WINNER architecture is the RRM Server. The RRM Server is communicating periodically with the associated BSs and takes the RTTMs so that at any given time it is able to know the state of each BS and if it is in a close to congestion situation. This communication between the RRM Server and the base stations can be either periodical and/or on demand, so when there is a need for RRM decision, the RRM Server can ask the BSs about the RTTMs. When the communication is periodical, the RRM decisions by the RRM Server are very fast, since there is no delay in demanding the BSs for the RTTMs. On the other hand, the decisions could not be the nest, since the RTTMs that the RRM Server has stored for the BSs could be very “old” and the state of the BSs could have changed. This can be solved by a frequent periodical communication between the RRM Server and the BSs, which can (on the other hand) increase the signalling through the network. The centralised version of the handover process has the main idea that the RRM decisions are taken by the RRM Server. This node is the one that gets the measurements from the BSs and makes the list of candidate for handover BSs and communicates with the target BSs for the completion of the handover process.

The distributed handover process has the following steps:

- HO trigger and initiation either from BS or from UT (or the HIS)
- Current BS communicates with neighbour BS (through IBB) and gets their RTTMs
- Current BS makes HO decision about target BS
- Current BS communicates with target BS (through IBB) about the HO completion

Figure 4-29 depicts the distributed handover signalling.

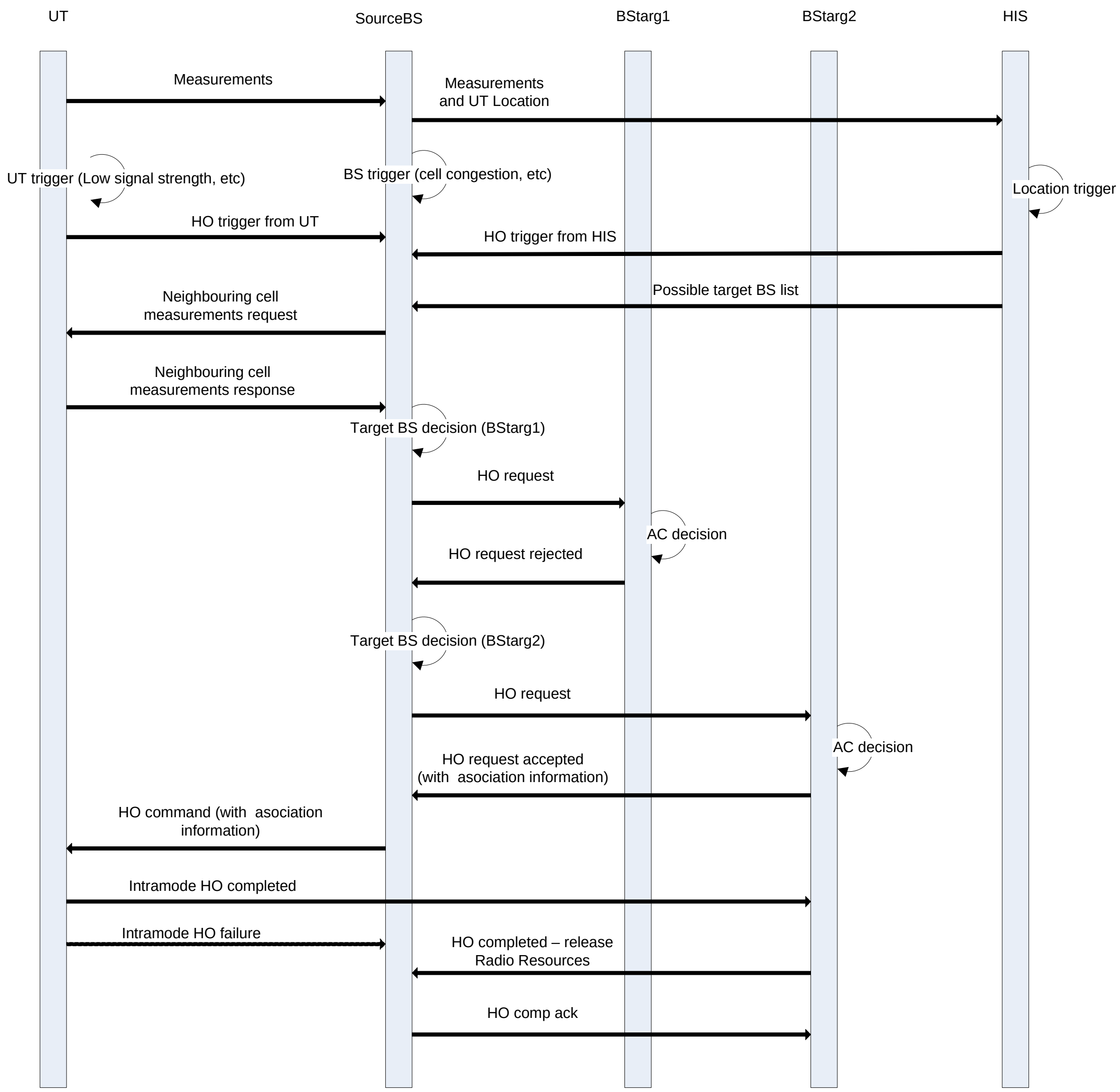

Figure 4-29: Distributed handover process

The centralised handover process has the following steps:

- HO initiation either from BS, UT or RRM Server
- RRM Server gets RTTMs from BS (or checks the measurements that already has)
- RRM Server makes HO decision about target BS
- RRM Server communicates with target BS about HO completion

Figure 4-30 depicts the centralized handover signalling.

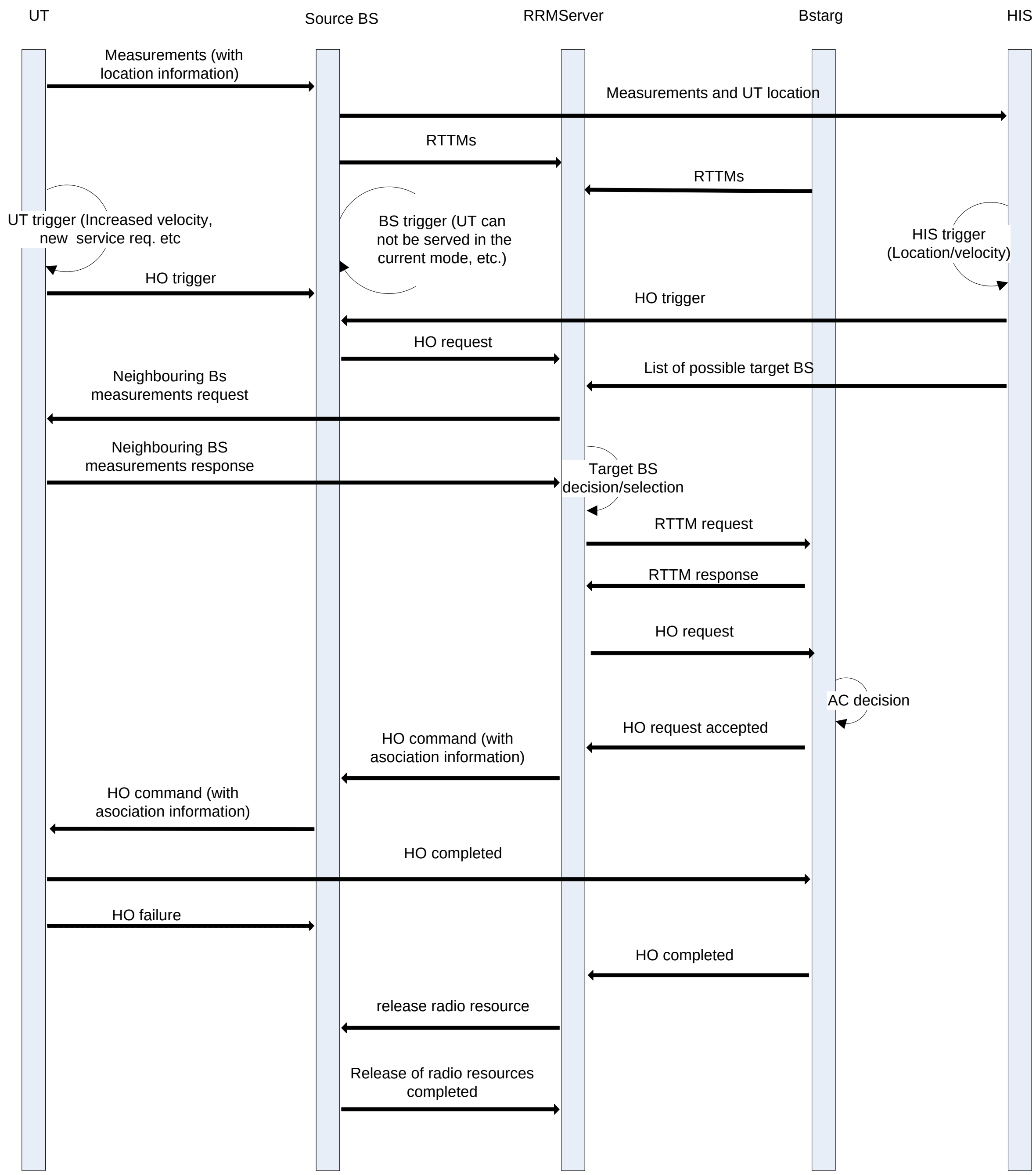

Figure 4-30: Centralized handover process

In order to exploit the advantages of the above described approached for the handover we have defined another approach that combines the above methods. In this hybrid solution, the RRM Server still exists, but it doesn't take the RRM decisions. The RRM Server still communicates with the BSs that it controls to get the RTTMs. The current BS communicates also with its neighbour BSs and gets their RTTMs. It also communicates with the RRM Server and forwards HO triggers to it so the RRM Server can execute the RRM algorithms and propose a candidate set of target BSs to the current BS to let it take a better decision. The current BS is the one that takes the HO decision and the RRM Server can assist to that. Also, there is another option: the current BS gives the RRM Server full authority to execute the handover. The trigger for the handover can come from the UT, the current BS, the HIS or the RRM Server.

Figure 4-31 show hybrid handover signalling process and, Figure 4-32 the hybrid handover process.

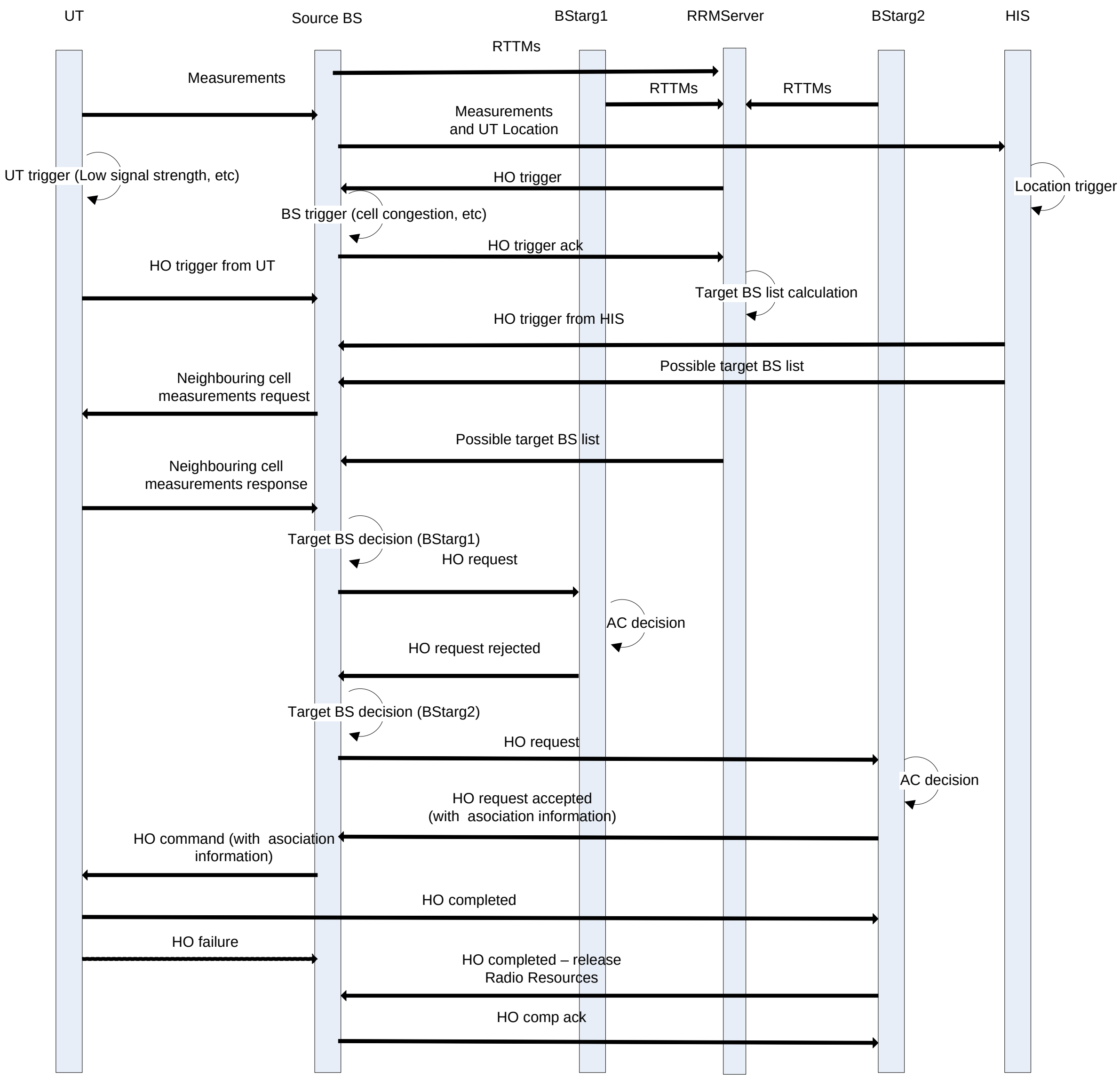

Figure 4-31: Hybrid handover signalling process

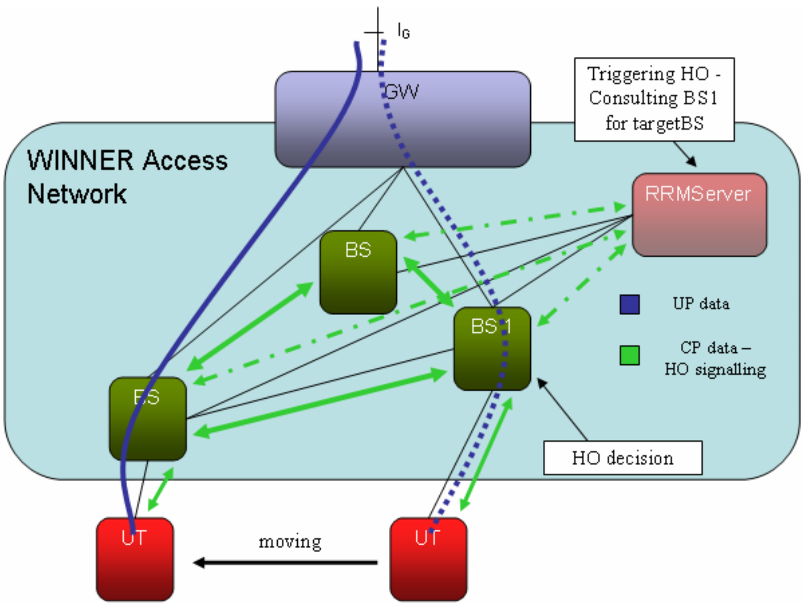

Figure 4-32: Hybrid handover process

4.2.3 Radio and IP handover

There are three different options for the handover related to the layers that are involved in the execution in the WINNER system. If the handover takes place only in Layer 2 then we have Radio handover. If the Layer 3 takes also place in the execution of the handover, then this could be IP or integrated IP and Radio.

Radio handover takes place when the UT, changing its attached radio access point, which could be a RN or a BS, mantains the same IP address after the completion of the handover. This seamless handover bases only on a switching process in the BS/GW. Since the UT has the same IP address after the HO process, the data flow is not interrupted since the application does not see any changes. Only the routing of the packets to the destination BS/RN is changed. This is the most common and simple type of handover, performed in the legacy networks such as GSM/GPRS and UMTS and even WLAN, when the source and the target Access Points belong to the same network domain. The radio handover will be the most common type of handover also in WINNER.

IP handover takes place when the UT changes its IP address after the completion of the handover process. This normally happens when the UT changes the gateway it is attached to: in this case, it enters a different IP domain and it changes its IP address. To implement IP handover (without interrupting or dropping data flows) it is necessary to utilize protocols as Mobile IP.

In WINNER the radio handover will be performed in the following cases:

- Inter-mode and intra-mode handover in WINNER if the BSs involved are controlled by a same Gateway (pool).
- Hybrid handover, where radio and IP handover are both performed, for instance, for GW balancing. However, in reality, IP handover can usually not be separated from L2 handover. On the one hand, an IP handover is triggered by a L2 trigger, on the other hand, there is frequently at the same time a L2 link down with an old base station and a new link up with a new base station. Therefore, most IP handovers require an implicit radio handover.
- Handover between WINNER and legacy networks if the different RATs are tightly or very tightly coupled. In this case, the different RATs are under control of a same Gateway (pool).

IP Handover can occur in the following cases:

- Load balancing between two Gateways within a same Gateway pool or in two different Gateway pools. In this case, user should change its anchor point to a new Gateway, therefore change its IP address and perform an IP handover.
- Handover between WINNER and legacy networks if the WINNER system and legacy networks are not tightly or very tightly coupled. In other words, if the different RATs are connected through gateway(s), then a handover between these heterogonous networks will be an IP handover.

The integrated handover case is when the ip and radio handover occur at the same time. The IP handover will be perfomed when the user changes GWs and this result in a change of the IP address. Radio handover also is performed when the user changes the BS to which the UT is connected and a new connection to a BS that is controlled by the other GW is established. The handover mixture is shown in Figure 4-33

The integrated handover can be supported both by a centralised or a distributed architecture. In a centralised architecture, where the decisions are taken by a central entity like the RRM Server, the handover mixture process would give very good solutions for the load balancing problem. The RRM server will have the measurements from the BS to directly trigger and control the radio handover to prevent call drops during non proper radio handover. The RRM Server will offer the candidate BS information to the source BS that retrieves the target GW address for the users needed to perform this type of handover and will send the appropriate commands to the UTs and the source BS and GWs to start the handover process. The handover mixture process would include signaling exchange between source and target GWs and source and target BSs directly.

If the network architecture is distributed, then it would be needed more signaling from the source GW to find a less loaded GW and a corresponding BS and this would increase the network signaling. The handover mixture would be exploited in maximum in a hybrid architecture, which would be a mixture of a centralised and a distributed architecture. In this hybrid architecture, the main decisions would be taken by the BSs, but there would be also a central RRM Server that would assist the BSs in their decisions. The RRM Server in that case would give advice to the BSs and when one BS cannot admit a user, then the RRM Server would come up in charge for finding the target GW-BS or the target network (in case of intersystem handover) for admitting the user. This architecture gives full power to the handover mixture since it can take very quick decisions by the distributed part and very good centralised decisions by the RRM Server advice.

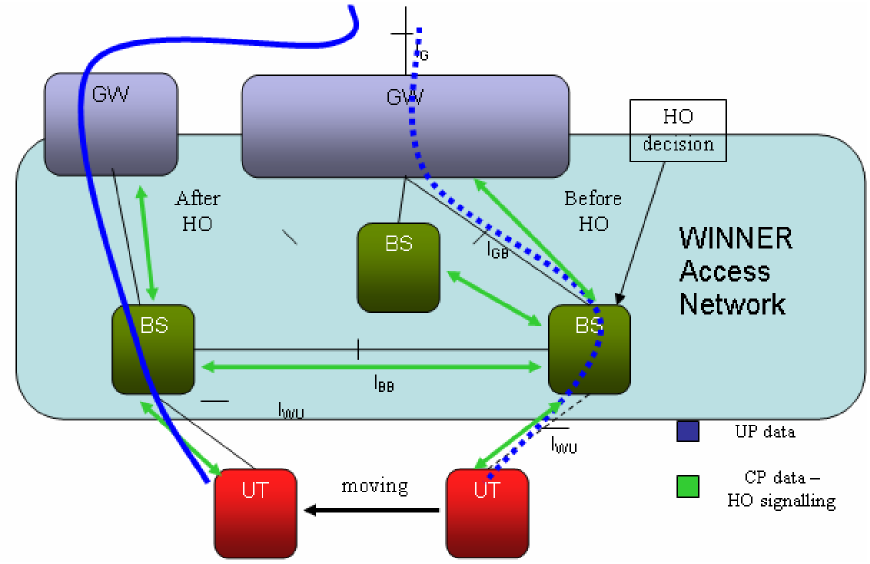

Figure 4-33: Integrated radio and IP handover

In Figure 4-34 is presented a more detailed description of the hybrid handover. The UT changes a BS so it can be considered as a radio handover, since the radio link has changed, but at the same time it changes a gateway, so it changes also the IP address. Since the radio and IP handover are performed at the same time, it is an integrated handover. The integrated handover can also be completed in two steps to avoid the delay of a simultaneous L" and L3 handover. In the first step the radio handover will be processed. This means that the UT will change the BS and will perform a radio handover to the new BS. This radio handover will be very fast and the process is shown in the first figure. The second step is the IP handover. After the UT has performed the radio handover to the new BS, then it performs an IP handover to the new GW, so it changes a GW and it enters a new IP domain, so it changes IP address and this is an IP handover. This second step is shown in the second figure.

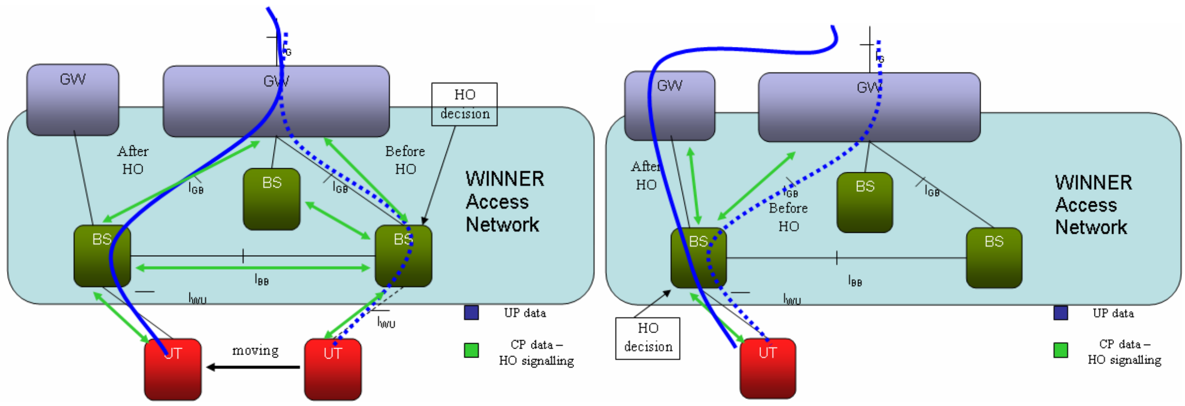

Figure 4-34: Integrated radio and IP handover, step 1 (Radio HO) step 2 (IP HO)

4.2.4 Load control

To avoid the use of IP handover, not as efficient as radio handover, the concept of the pool of gateways has been introduced. The pool of gateways will share the resources of the pool of base stations which they are connected. When there is a handover between BSs in the same pool, the GW normally will be not changed and thus the IP address will remain the same (see Figure 4-35). This suppose that the logical association between the UT and GW is tried to be preserved independently of the serving BSs.

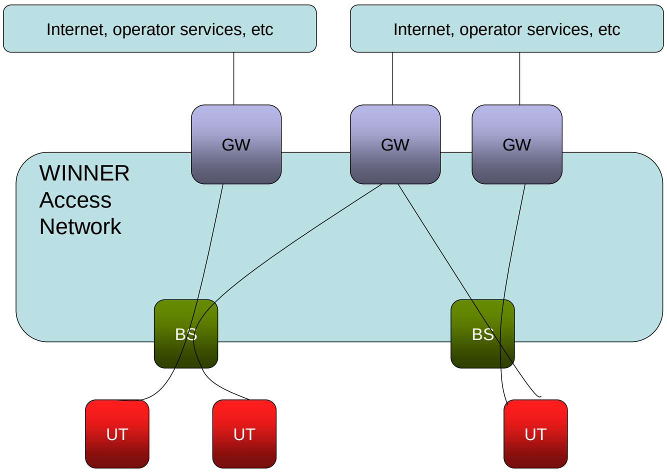

Figure 4-35: GWs as a pool of resources

Indeed, each GW may be associated to each BS in the pool area. Such an association avoids IP handover between GWs in the given pool area. This is shown in Figure 4-36.

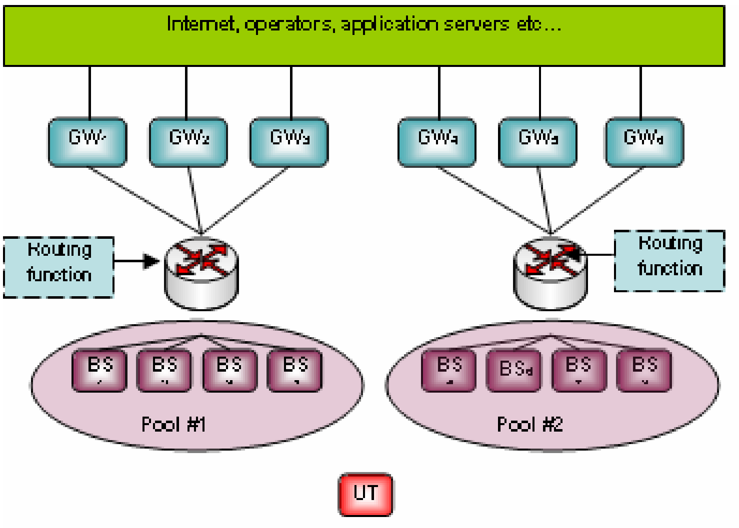

Figure 4-36: Pool area of GWs.

The traditional approach is a hierarchical structure, where each GW is associated with a set of BSs serving their own location area and providing a direct mapping between a GW and the area covered by associated BS. The GW is an anchor point for external routing, and also is bridge between the UT and the external world, outside WINNER RAN, through the IG interface. Moreover, the pool capacity is easily optimized by adding or removing GWs. Finally, it provides redundancy, that is, in the case of a GW failure, the users can be handed over to any other GWs in the pool, and at the same time load balancing between GWs can easily be achieved.

Therefore, the GW association will be preserved even when a user handovers to a BS that is controlled by a different GW, belonging to the same pool. In that case, a change of the IP address is not necessary. This is the basic idea of the pool of gateways that are connected to the BSs through a routing function that enables the process described above and shown in Figure 4-36.

Load balancing between gateways suppose an IP handover between GWs that belong to the same or different pools. Two cases of load balancing are presented in Figure 4-37 and Figure 4-38.

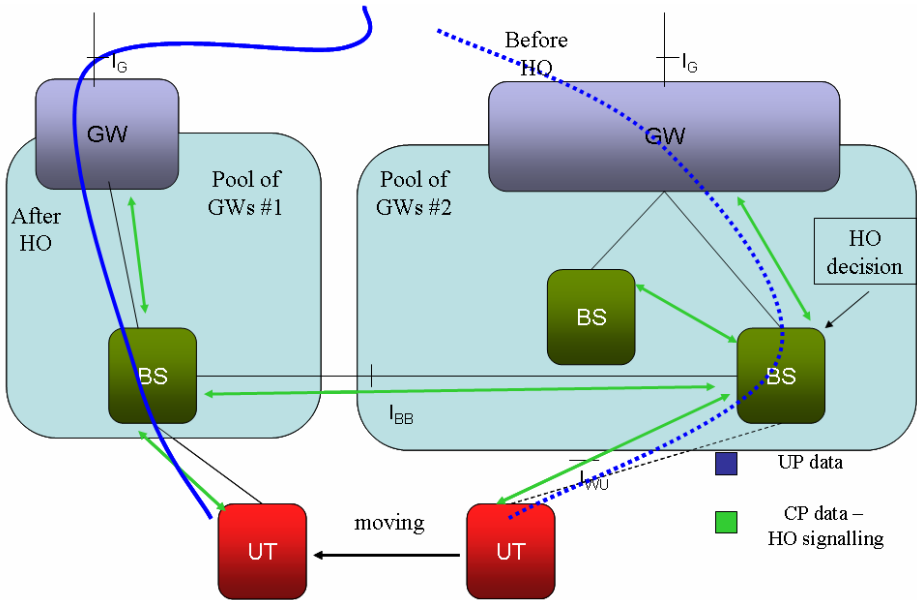

Figure 4-37: Load balancing case1 – different pools of gateways

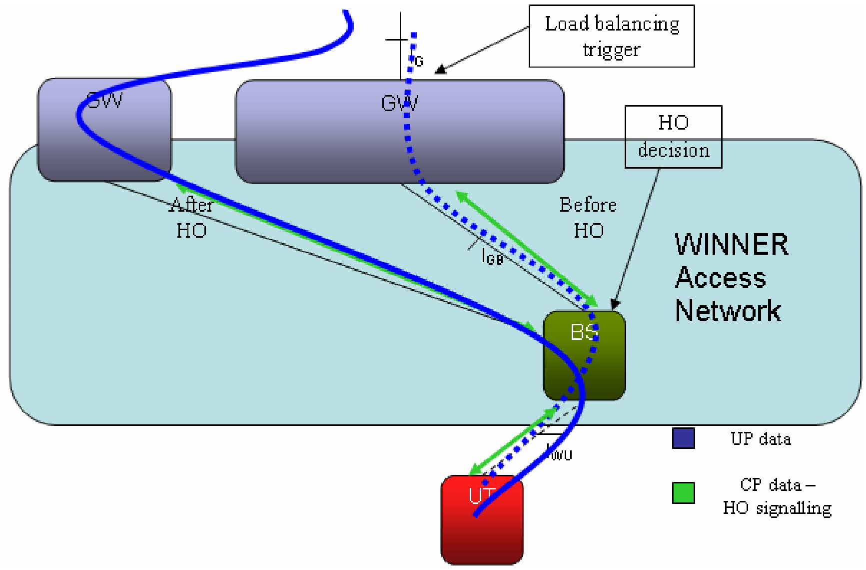

Figure 4-38: Load balancing case 2 – same pool of gateways

Load balancing between GWs is needed when a GW is congested or in order to achieve balanced distribution of load between the GWs. A GW can be congested when either there are too many active users associated to it or there are highly loaded users associated to it and the total load exceeds a predefined threshold in relationship to the GW capabilities. Then there would be a need to decongest the GW. Decongestion of the GW can be performed by decreasing the quality of service (QoS) level of the associated users or drop some of the associated users, or by the handover of a number of the associated users to another less loaded GW. The load balancing between GWs is another option for preventing congestion situations in the WINNER network, since by that the network load is being distributed to all the gateways and the BSs of the network.

4.2.5 Policy based mobility management

The policy-based mobility management scheme proposed for the WINNER system is based on previously developed cooperation radio resource management (RRM) algorithms including mobility management, for the successful interworking of WINNER with new and legacy systems. The rules that are applied for the support of mobility within the WINNER scenarios solve problems such as: who and what can access which resources and

on which network, what is the highest and lowest priority traffic, what are the levels in between, and how are they differentiated, how do we guarantee delivery of highest priority traffic (e.g., real-time applications), how can we guarantee the required for the delivery bandwidth, when and which traffic can we discard? In addition, based on the QoS demand, location determination can also be scaled into different levels by using different positioning methods. Further, the policy-based mobility management scheme includes actions related to RAT association, user and flow context transfer, handover decision, and deployment priority.
Mobility in the WINNER RAN is supported by traffic and control signalling from the user terminal (UT) to the base station (BS) that the UT is connected to, and also by the BS to BS control signalling. To ensure flexibility of the architecture, logical functionalities of the physical entities can be grouped according to the situation.
For example, in a cellular deployment, fast handover without service interruption is needed. For this purpose, the gateway functionality is introduced and a layer 2 tunnelling protocol provides fast handover. A WINNER specific physical node is needed in this case.
To enable efficient signaling and management between network and UT, all profiles are captured in the home operator domain for all the registered UTs. The RAN is responsible for enforcement of the policy determined by the core network.

The policy management is distributed between the home subscriber server (HSS) and the WINNER gateway. The RRM functions are based on the policy defined or previously agreed between the end subscriber and the WINNER operator and applied to execute the following actions:
- Radio Access Technology association;
- Flow establishment and QoS class setting;
- Handover priority setting;
- Context transfer in the U-Plane;
- Location determination;
- Deployment priority.

Mobility management therefore, should be implemented such that it will include aspects related to handling the session/application; load balancing; multi-homing; security; and all sides of mobility, i.e., endpoints, sessions, flows, interfaces, network/groups, flexible spectrum utilization, group and signaling managements.

The context of a UT is established mostly in the GW after its success of initial access. It is conceptually separated from the flow contexts of this UT to enable individual routing of these flows through different cells. The user context might be kept centrally in the RAN or moved from BS to BS in the case of a handover. The user context contains all information related to a UT in the system: user-id; flows associated to the user; physical context information, (e.g., location, direction of movement, speed).

The context of a flow is established and released by the flow establishment/release function. If a flow should be moved from one BS to another, the context of the flow has to be transferred to the new BS. This might happen in the case of radio handover due to a changed radio link condition or if multiple radio links are available to the UT (overlay cells) due to the shift of load between the cells. The transfer could happen within a one domain (micro-mobility) or between two domains (macro-mobility)

The WINNER air interface is optimized for the transfer of IP based services and the overlaying protocols (UDP, TCP, RTP, SIP). Since these protocols have been defined for the fixed line Internet where bandwidth limitation is not a severe problem, significant amount of the available capacity on the air interface would be spent for the transmission of the protocol header.
In a practical example the first packets of the flow are used for creating a context on both sides that contains information e.g. about constant or sequentially incremented protocol header fields. These packets are sent uncompressed. After the context is created the compressor compresses packets based on the context information. Uncompressed packets are sent from time to time and in the case of error recovery by the compressor to revert to normal operation and to reconstruct the context.

There is another problem in the use of wired TCP in a wireless environment, with eventual packet losses due to poor radio conditions that are interpreted by TCP protocol as network congestion, reducing significantly the data rate and provoking duplicate delivery of packets that however are buffered in the BSs. The reduction of performance could be sensible in the case of handover, an in general in circumstance with packet losses. To cope with this problem a proposal for the link layer segment labeling has been proposed in WINNER to improve the performance.

According to the service profile the user registered, direction of handover can be controlled by the policy. For instance, user A and B are both subscribers. A subscribes to lower class; on the contrary B is rather a premium user. In that case, ranking of the handover candidate cell list can be differently ordered, e.g. user A has rank: LA cell x → LA cell y → WA cell z; however user B is allowed to set handover list with ranking WA cell z → LA cell x –LA cell y. Context of A has also higher priorities than B to be transferred between neighbouring cells. The policy defined HO priority setting helps the QoS guarantee for super class users.

4.2.6 RLC PDU context transfer

RLC context transfer of RLC PDUs in addition to RLC SDUs may help to reduce delays due to handover. With RLC PDU context transfer, if RLC SDUs are divided into many RLC PDUs of smaller size and only few of the RLC PDUs that belong to one and the same RLC SDU are lost, then it is not necessary to retransmit the whole RLC SDU (as it is if only context transfer of RLC SDUs are performed). RLC PDU context transfer may also be efficient even if there is a one-to-one mapping between RLC SDUs and RLC PDUs. Assume that the receiver has successfully received a number of RLC PDUs before handover, but the sender has not received any status information, which could occur if a poll request or status message is lost, or if the poll interval is long. If only RLC SDU context transfer is applied, then all unacknowledged RLC SDUs need to be retransmitted after handover even though it is unnecessary, since all data were successfully received already before handover. With RLC PDU context transfer, on the other hand, the sender could poll the receiver after handover and get status information about the successfully received data. Thus, unnecessary retransmissions could be avoided.

One of the more important questions regarding efficient RLC operation is how often poll requests and status messages should be exchanged. Too frequent transmissions waste radio resources, increase power usage, and may even trigger unnecessary retransmissions [Alcaraz06]. Too infrequent transmissions may instead stall the RLC transmission window, which may result in under utilisation of the radio link. In 3GPP RLC [3GPP322], a multitude of options for polling are specified, some of which are one-shot (when some condition becomes true, e.g., last PDU in buffer) and others which are recurrent (expiry of poll timer and periodic polling). Incorrect configuration may cause deadlock. Recurrent polling is required to avoid deadlock [Chen03]. In order to avoid too frequent transmissions, a poll prohibit timer and a status report prohibit timer can be used. A poll prohibit timer sets the limit for the minimum interval that is allowed between two poll requests, and a status report prohibit timer between two status reports.

In the following example, we have chosen to represent polling with a poll timer. The amount of outstanding data that is buffered in the source BS at handover depends on the available bandwidth, the delay and the setting of the poll timer,

$$b(T + d + d)$$

where *b* is the available bandwidth, *d* is the one way delay over the radio link, and T is the timeout value of the poll timer. The amount of outstanding data is the product between the bandwidth and the time required to transmit the data, which is the timeout value, T, and the time it takes to transmit a poll request and to get a status message back, $d + d$.

In the following, it is assumed, that all transmitted data are successfully received before handover, but that the status message has not reached the sender. Furthermore, it is assumed that there are always enough data available to fully utilise the radio link, and that the RLC transmission window is large enough not to get stalled.

If only RLC SDU context transfer is applied, then RLC state is reset and all forwarded RLC SDUs will be transmitted again to the UT, even though they were successfully received before handover. The time required to retransmit all outstanding data on the new path may be longer or shorter than the time to transmit the data over the old path, depending on the delay on the new path. In case of BS_i-BS_jRN handover, for example, the delay would be longer on the new path, and in case of BS_iRN-BS_j it would be shorter. If the delay is the same as before, then it will take T to retransmit the data and an additional $d + d$ to exchange poll request and status message. In total it takes $T + d + d$ before the RLC sender receives the status message about successful transmission. The delay may be reduced and unnecessary retransmission may be avoided, if RLC PDU context transfer is used and a status message is transmitted immediately after handover is completed. If the status message reaches the RLC sender, no retransmission is needed, since all outstanding data were successfully received already before handover. With RLC PDU context transfer, the RLC sender will receive the status message about successful transmission after d instead, which is a shorter delay than with RLC SDU context transfer.

In Figure 4-39 the outstanding data vs. poll timer is shown, for some example combinations of bandwidth and delay. A delay of 2 ms could illustrate the case where a UT communicates directly with a BS, and a delay of 20 ms communication via a RN (including queuing delay in the RN). With a poll timer set to 100 ms, b = 10 Mbps, and d = 20 ms, for example, the outstanding data is 1400 kbits. With only RLC SDU context transfer, transfer of new data could start first after 140 ms, after the forwarded data are retransmitted (provided that the new path has the same characteristics). With RLC PDU context transfer, and assuming that a status message is transmitted immediately after handover, transfer of new data can start after 20 ms. If only RLC SDU context transfer is applied, then the extra delay caused by the unnecessary retransmission of RLC SDUs may have a negative impact on the performance of higher layer protocols. New data is delayed, if the buffered RLC SDUs are retransmitted first. Depending on the relation between the delay and the TCP retransmission timer, this may lead to TCP retransmission timeout and that data will be unnecessarily retransmitted also by TCP. The transmission rate would be drastically reduced in response to the timeout. User experience could suffer, especially in cases when TCP is used for real time applications. TCP can be used to transmit real time data, such as video streaming and voice (e.g., Windows Media, Skype), if UDP connectivity is limited by firewalls. Also applications that use UDP may suffer if RLC is used in acknowledged mode, as for the video streaming application in [Lo05]. Low delays are desirable for video streaming, even if delays can be compensated for by adjusting the playback buffer.

We proposed RLC PDU context transfer as a value added function in WINNER. RLC context transfer of both SDUs and PDUs could apply the following scheme:

In DL

1. Before handover, the RLC receiver in the UT transmits a status message to the source BS.
2. During handover, before it is completed, the RLC sender in the source BS forwards buffered RLC SDUs and RLC PDUs in the transmission queue to the target BS. Also RLC state is transferred.
3. After handover, the UT transmits a status message to the target BS, and the forwarded RLC PDUs that have not yet successfully received by the UT are transmitted.

In UL

1. Before handover, the RLC receiver in the source BS transmits all successfully received RLC SDUs to IPCL in the GW and any remaining RLC PDUs to the target BS. Also RLC state is transferred.
2. After handover, the target BS transmits a status message to the UT. The UT continues to transmit RLC SDUs and unacknowledged RLC PDUs.

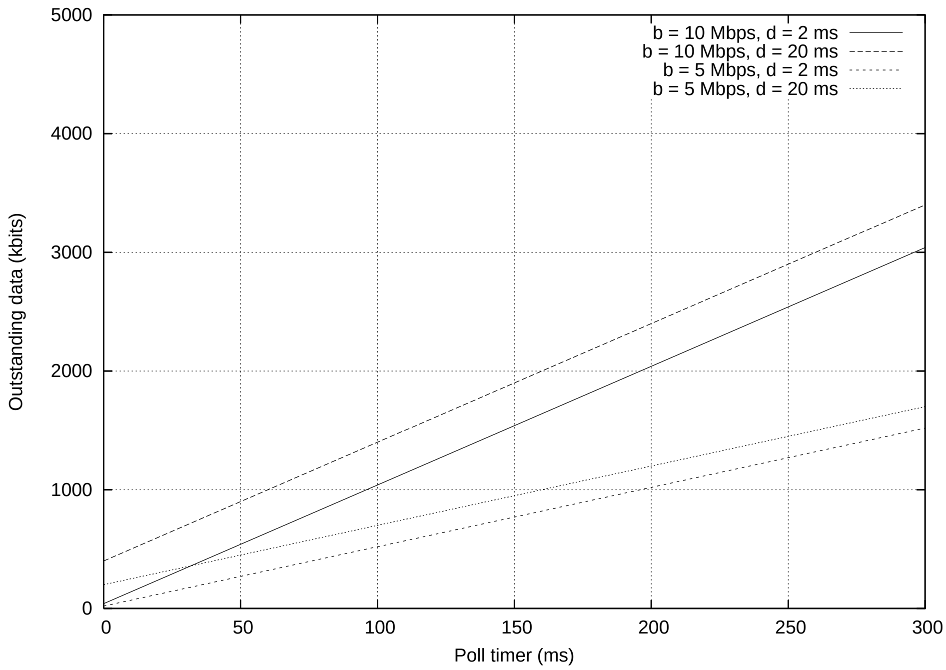

Figure 4-39: Outstanding data vs. poll timer

Just as for RLC SDU context transfer, inconsistent states shoud be avoided if possible. Therefore, we propose that, in DL, the RLC receiver in the UT is instructed from higher layers to transmit a status message to the source BS before handover. In UL, this seems less important, since the RLC sender remains in the same node, the UT, after handover. We also propose that higher layers instruct the RLC receivers to transmit a status message after handover. If the sender fails to get status before handover, unnecessary retransmission could be avoided with a status message after handover.

4.2.7 Fuzzy Logic based handover

Fuzzy logic is another added value for WINNER handover. Thanks to it, the WINNER handover strategy can take imprecise or incomplete or noisy measurements as inputs. Another advantage of using fuzzy logic is that numeous inputs can be integrated into several important inputs, such that handover decisions can be made much easier. The fuzzy system uses a list of rules to control the system, and system behaviour can be tuned, simply, by modifying the appropriate rules. In a same way, new inputs can also be easily introduced into the WINNER system. The fuzzy logic based handover scheme is shown in Figure 4-40.

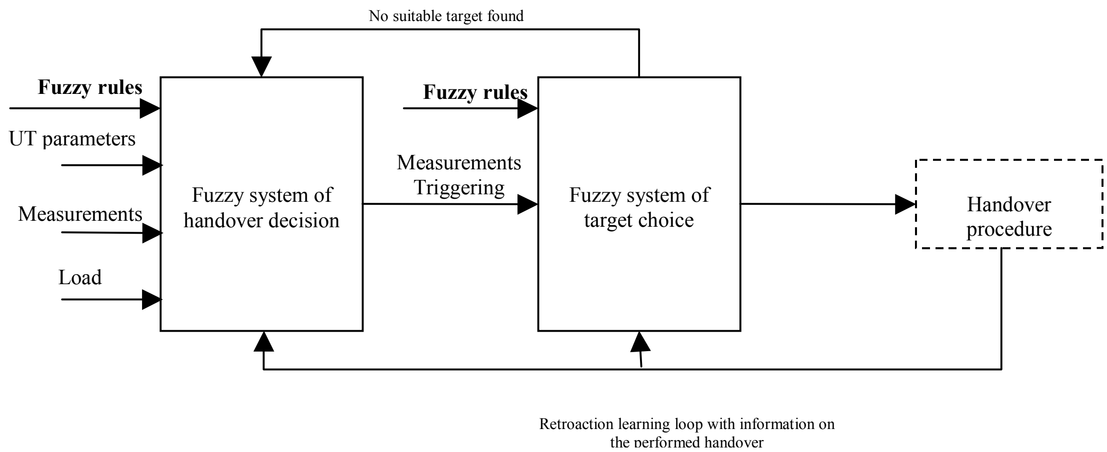

Figure 4-40: Fuzzy logic based handover scheme

Fuzzy logic is quite adapted to radio resources management because of radio environment fluctuations and uncertainty (measurements averaging, shadowing, traffic model …). Thus, it can be used for inter-mode and intra-mode handover decisions. Furthurmore, thanks to fuzzification, it is possible to compare quantities from heterogeneous RANs. Therefore it is also helpful for inter-system handovers. In complex systems, fuzzy models give more easily assimilated information than precise models. We also introduce learning techniques which can furthur improve the fuzzy system parameters and enhances its performances accordingly [WIND482].

The fuzzy logic based inter-system handover simulation results are presented in Section 6.4. Those results show that the fuzzy logic can effectively improve the inter-system handover efficiency, and then increases the system performance in terms of application throughput and response delay. Because that fuzzy logic itself is not dependent to any radio system, those results can even be generalized to intra- and inter-mode handover, and we can then conclude that fuzzy logic can be benefic for WINNER system.

The decision process can be assisted further by capturing the specific application requirements in terms of bandwidth, delay, and packet loss that also relate to the overall QoI the user expects. Different threshold requirements can be devised and quickly evaluated by means of the inference and learning techniques implemented at the convergence layer and these results can be made available at the lower layers (e.g., link layer). For example, to support better a real time video streaming application, the delay between each packet received can be monitored and a handover can be triggered if the delay variance (jitter) goes beyond a pre-defined threshold.

In this context, fuzzy rules can be designed to assist the decision making process during admission control [Mihovska07]. Admission control should explore a set of measured parameters to make the decision of accepting or rejecting a requesting call. This type of control scheme makes little or no allowance for measurement uncertainties. However, in a wireless system, due to user mobility and varying channel conditions, the measurements obtained are, in general, not accurate.

Table 4-3 describes the fuzzy rules applied during admission control for the purpose of congestion detection or congestion resolution. Each service request (in the WINNER network) will have to go through the AC process.

	No Congestion	Congestion
Throughput < Capacity threshold AND Load < Congestion Threshold	Increase Bit rate starting from high priority session/user to the lowest one *	CONGESTION RESOLUTION PROCESS
Throughput = Capacity threshold	Noop	CONGESTION RESOLUTION PROCESS

Table 4-3: Fuzzy rules for admission control to assist congestion detection and congestion resolution

If *No Congestion* **AND** *System Throughput < Capacity* **AND** *Load < Congestion Threshold* **THEN** *Do Bit Rate Increment until Throughput = Capacity*

In the following the advantage of the fuzzy rules for admission control are investigated by means of a theoretical model. .
Several assumptions and limitations are made.

1. Only the session, i.e. flow context, is considered without caring about the user context, such as its priority, type of contract, etc.
2. Mobility is not considered
3. Five cells are considered, i.e. Winner LA, Winner MA, Winner WA, UMTS, and WLAN
4. The cells are considered to be overlapping with each other
5. Four traffic classes are considered:
   a. VoIP (real time and low B/W)
   b. Web browsing (non-real time and low B/W)
   c. Video streaming (real time and high B/W)
   d. File transfers (non-real time and high B/W)

6. Each traffic class has different priority where traffic class (a) is the highest while (d) is the lowest priority. Furthermore, each of them has different characteristics in terms of minimum delay and required rate, packet size, session arrival rate, and duration. They are described in the following table [Kim06] [Soladni03]:

15. Session arrival rate is defined for each cell. For the WINNER RAN, three traffic load scenarios (TLS) are defined (see [WIND6.12.1]), i.e. low, medium, and heavy, while for UMTS and WLAN the TLS are assumed to be all low arrival rate. Furthermore, they are all assumed to follow the Poisson distribution. This is mentioned in Table 4-4 and Table 4-5 [Kim06]:

	Video streaming	VoIP	File transfer	Web browsing
Minimum delay (s)	0.08	0.15	0.5	1
Required rate (kbps)	500	32	512	64
Packet size (bytes)	1500	70	512	512
Duration (s)	Exponentially distributed µ:150 s, min:10 s, max: 600s	Exponentially distributed µ: 120 s	File size: uniformly distributed, min: 10 kB, max: 70MB, overhead: 24 bytes $duration = \frac{file\_size}{rate}$	1 s

Table 4-4: Characteristics of the traffic classes

Session arrival rate (sessions/second)	Video streaming	VoIP	File transfer	Web browsing
WINNER Low	0.041	0.1	0.041	9.23
WINNER Medium	0.059	0.3	0.059	13.185
WINNER Heavy	0.076	0.6	0.076	17.141
UMTS	0.02	0.2	0.02	8
WLAN	0.04	0.2	0.03	9

Table 4-5: Session Arrival rate for the WINNER TLS

16. The capacity for each cell is defined in Table 4-6. The capacity values for the WINNER network are actually the maximum aggregated data rate for DL and UL by assuming the highest modulation level and coding rate which is 64 QAM with ¾ coding rate. It is also assumed several number of chunks per frame and sysmbols per chunk, which are 460 chunks/frame and 96 symbols/chunk for LA and MA deployment, and 288 chunks/frame and 120 symbols/chunk for WA deployment (WIN2D6137).

	WINNER LA	WINNER MA	WINNER WA	UMTS	WLAN
Capacity	359.375 Mbps	359.375 Mbps	180 Mbps	10 Mbps	54 Mbps

Table 4-6: Capacity for each WINNER Cell

9. The input parameters of HO initiation are delay (D) and offered traffic (T). These are calculated for each cell and for each traffic class, based on the formula in [WIND 6.2]. How they are calculated is explained as follows:

Offered traffic per traffic class: $T_n = \sum P \cdot \mu \cdot r$ (Eq. 1) where:

P is mean session arrival rate, μ is mean duration, and r is mean rate.

Packet arrival rate per traffic class: $\lambda = \frac{T}{s_n}$ (Eq. 2) where s is packet size.

Delay per traffic class: $D_n(C) = \frac{\sum_{i=1}^{N_{ps}} \lambda_i s_i^{(2)}}{2\left(C - \sum_{i=1}^{n} \lambda_i s_i\right)\left(C - \sum_{i=1}^{n-1} \lambda_i s_i\right)} + \frac{s_n}{C}$ (Eq. 3) where C is cell capacity.

10. The following condition must be fulfilled: $D_{n\_required} < D_{n\_measured}$ (Eq. 4) and $\sum_{i=1}^{n} \lambda_i s_i < C_n$ or $\sum_{i=1}^{n} T_i < C_n$ (Eq. 5), otherwise the HO initiation will be triggered

11. In order to choose the most suitable cell after the HO process, the measurement matrix from different cells and the traffic class requirements are used. It will be done by applying Fuzzy Multiple Attribute Decision Making (MADM) [Zhang04]
12. Observation points are the impact different TLS and the frequency of monitoring, i.e. how frequent the network parameters are calculated.

13. Performance matrix: throughput, delay, rejected users, dropped users, etc….

Figure 4-41 shows the flow chart for the model described above.

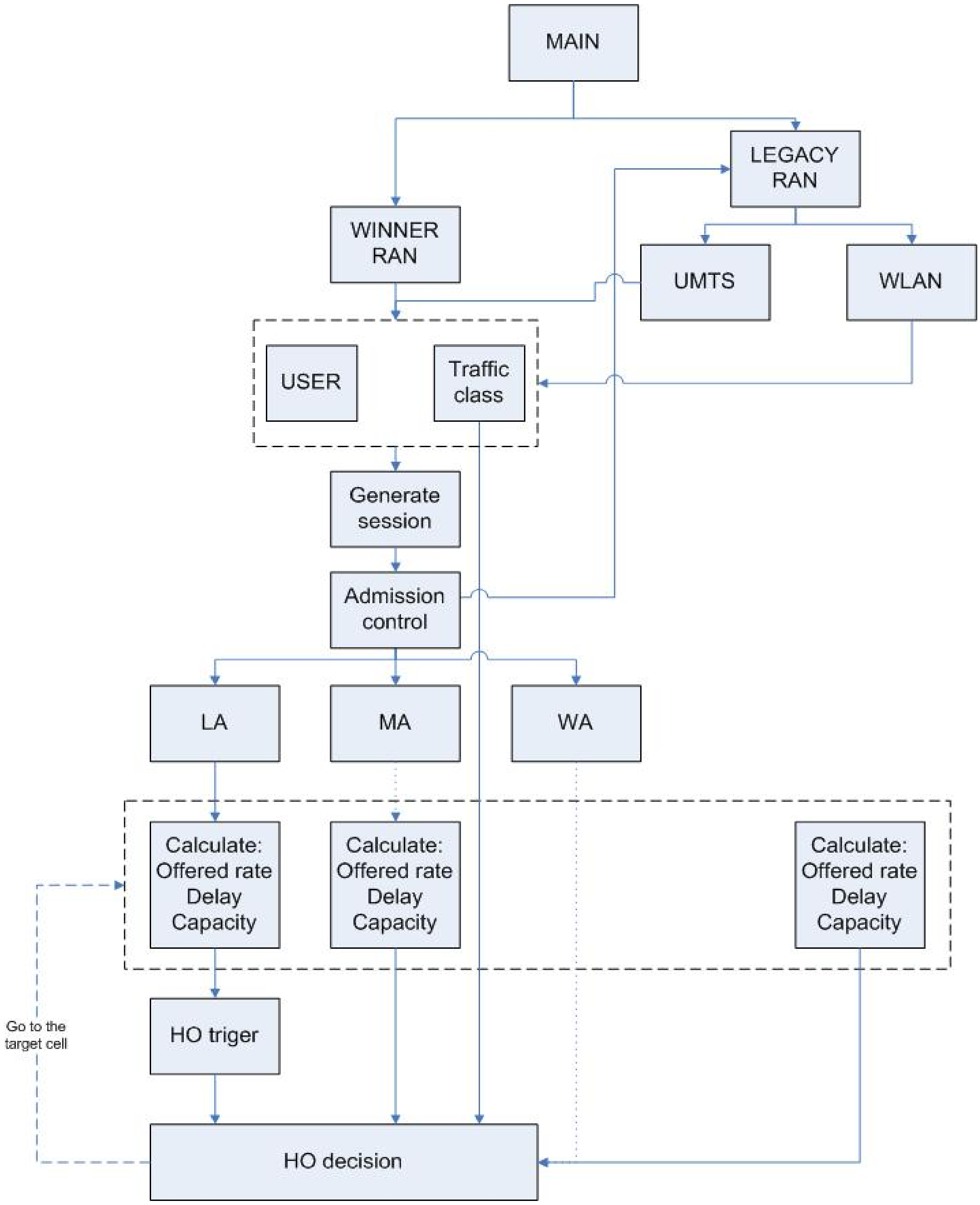

Figure 4-41: Flow chart for the fuzzy logic admission control model

The different blocks of Figure 4-41 are described below.

- **Admission Control Block**:

  Each service request (in WINNER network) will have to go through this process. The session will be served by the suitable BS according to its traffic class, otherwise if the desired BS cannot support, i.e. due to overload or something else, other supportable WINNER mode will be chosen, if it still not possible, ISHO will be carried on.

- **CELL Block**:

  Each cell will have its own parameters, i.e. delay and offered traffic, and also the number of active sessions and will be divided according to traffic class within that particular cell, which is described in Figure 4-42

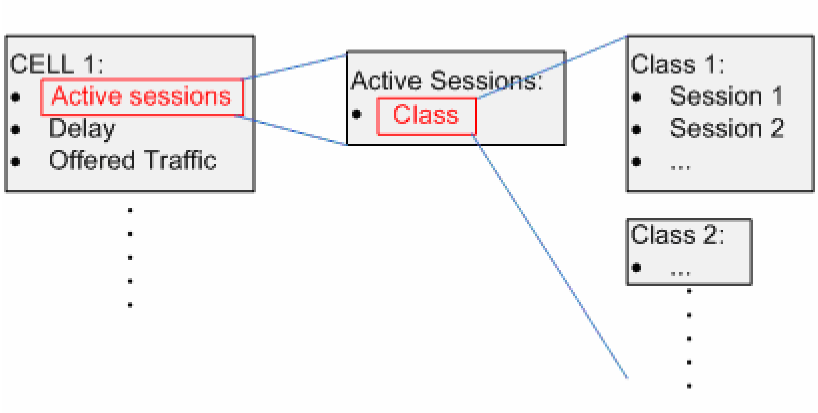

Figure 4-42: CELL block structure.

- **TRAFFIC CLASS Block**:

  It defines the minimum delay, required rate, packet size, default winner mode, and the arrival rate.
- **Generate Session Block**:

  The session arrival rate is defined differently for each traffic class and cell according to already defined parameters in the traffic class block.
- **CALCULATION Block**:

  It is already explained in the previous section.
- **HO Trigger Block**:

  This block is applied for each cell and traffic class, i.e. there will be 4 fuzzy systems within one cell for each traffic class. It uses fuzzy system with two inputs, *offered traffic* and *delay*, and the output is HO trigger. Each input has 3 triangle shape membership functions (MF), i.e. low, medium, and high. The membership values vary from zero to the *maximum cell capacity* for the *offered traffic*, and *maximum required delay* of certain traffic class for the *delay*, while the membership degrees vary from 0 to 1. The HO trigger output has 4 triangle shapes MF, i.e. No, Probably No (PN), Probably Yes (PY), and Yes. The MF values are normalized, i.e. from 0 to 1, and the degrees also from 0 to 1.

  The fuzzy rules of HO trigger are described in Table 4-7:

Traffic/Delay	Low	Medium	High
Low	No	PN	PY
Medium	No	PY	Yes
High	PN	PY	Yes

Table 4-7: Fuzzy logic rules applied to handover

If the output of the *HO Trigger* is more than 0.5, the system will go to the *HO Decision*, i.e. to choose the suitable cell for HO, otherwise there is no HO. Later on, the output of *HO Trigger* will be used as the input for the *HO Accept*.

- **HO Decision Block**:

  There are some proposals for the Fuzzy MADM based HO [Zhang04] [Guo06] [Navarro06][Song05]. The differences among them are in the way of weighting different input variables. According to [Guo06], there are several algorithms of Fuzzy MADM based HO in heterogeneous wireless network, i.e. MEW (Multiplicative Exponent Weighting), SAW (Simple Additive Weighting), TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), and GRA (Grey Relational Analysis). The description of SAW can also be found in [Guo06],with the more details explanation on getting the so called *weighting vector*. In this work, SAW algorithm from [Guo06], is considered for the simplicity. The parameters to be evaluated in are offered traffic, delay, and mean arrival rate.

- **HO Accept Block**

  This block is meant to evaluate if the selected RAN is able to accept the HO session with the required QoS. It is also applied for each cell and traffic class. The inputs are the *HO trigger* of the *current cell*, and *offered traffic* and *delay* of the *target cell*. The output of this will be “Yes” of “No”. If the output is “Yes” then the session HO to the selected cell is carried out, otherwise it will stay at the current cell.

4.3 Location service support

4.3.1 Description

The location determination system function handles location requests, originated either from the UT or from some service in the network. When a location request is received, the location service support makes sure that user’s privacy is not violated and/or checks that using the location service is allowed and charged for the specific user, and then it initiates the positioning process. Generally, the location determination (LD) can be done within the UT using performed measurements and information sent by the location service support function, or within the location service support using measurement reports sent by the UT and/or BSs and relay nodes involved in the location estimation process. If appropriate capabilities of the UT are available, also global navigation satellite system (GNSS) based support from GPS and/or Galileo can be included.

4.3.2 Requirements

According to the final WINNER system requirements in D6.11.4 (R8) “WINNER System will provide stand-alone location information that can be used as input for user and system-side applications, e.g., emergency services” [WIND6114]. D6.11.4 (R7) further says that “WINNER System shall aim at achieving user emergency call requirements”.

Concerning accuracy requirements “WINNER shall at least meet the requirements for the unified E-112 system agreed by the EU” (R26). However, currently no agreed European requirements exist, just recommendations by [CGA02]. The requirement "WINNER System will maintain compatibility with existing location information mechanisms (GPS, Galileo)" (R9) concerns an “add-on” for a better LD performance for, e.g., location based services or more advanced handover.

4.3.3 Measurements

In [WIND481] general positioning methods were introduced and described in a more abstract way. Descriptions were given for

- Time of arrival (TOA);
- Time difference of arrival (TDOA);
- Received signal strength (RSS);
- Angle of arrival (AOA);
- Cell ID.

However, different QoS requirements for the different applications require the flexible and scalable RAN as it is proposed by WINNER. Applications that can use the location of the UT could be: point-to-point navigation, emergency call handling, location-based handover, or location-based service provisioning. In order to cater for different QoS demands, the location information should also be scalable in order to provide appropriate accuracy for the involved UTs with different profiles [WIND482].

Generally, the location estimation can be done within the UT using measurements and information sent by the location service support function, or within the location service support using measurements sent by the UT and/or BSs and relay nodes involved in the location estimation process. The best stand-alone based performance can be obtained by including timing measurements of the UT. In-band timing measurements in cellular network based positioning are usually based on TDOA measurements [WIND482]. They are based on the idea to find the starting point (TOA) of the incident OFDM signals to estimate distances between the UT and the BSs using the included pilot sequences [MTL+07]. If the GNSS based positioning information is included (GPS, Galileo), the performance can be further improved [WIND482].

If the positioning is completely done in the UT (UT based mode), information about the positions of the BSs has to be available at the UT. Also the time-offset of the BSs in non-synchronized networks has to be measured or communicated. If the LD is done in UT assisted mode, the measured TDOAs have to be communicated to the network which calculates the final position estimate.

4.3.4 Location determination

In [WIND482] approaches to obtain TDOA timing information by OFDM synchronization algorithms were presented. Furthermore, the integration of GNSS based positioning techniques was described for the static case. However, for further improvement we can include tracking algorithms for the solution of the navigation equation in the dynamic case.

Usually, the user with its UT is moving around a certain track in different scenarios. Clearly, there are certain correlations between the positions of the UT over time. This information can be integrated in the overall position estimation process and can help to improve the estimates in average.

The Kalman filter (KF) is a flexible tool for providing such positioning estimates in the context of tracking applications. However, the standard KF just performs optimal if several criterions on, e.g., linearity or Gaussianity, are fulfilled which is usually not the case in practical applications. Nevertheless, even if these conditions are not fulfilled completely, the KF gives reliable and robust estimates.

Generally, the KF is a generalization of the Wiener filter, where the restriction of the Wiener filter that signal and noise are stationary is not mandatory. It is a sequential minimum mean square error (MMSE) estimator of signals embedded in noise, where the signal is characterized by a dynamical or state space model which for the here considered case is the mobility model. In the context of positioning by a mobility model the character of the user's movement over time can be described. If signal and noise are jointly Gaussian, the KF is the optimum MMSE estimator [Kay93].

4.3.4.1 Kalman filter

The scalar Kalman filter case can simply be extended to the vector case straightforwardly yielding the so-called vector KF which is considered here [Kay93]. In this case, the state space and observation equations can be written as

$$\mathbf{s}[n] = \mathbf{A}\mathbf{s}[n-1] + \mathbf{u}[n] \in \mathrm{R}^{p} \tag{4.1}$$

and

$$\mathbf{x}[n] = \mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] \in \mathrm{R}^{q}. \tag{4.2}$$

The noise vectors for the state space equation $\mathbf{u}[n] \in \mathrm{R}^{p}$ and for the observation equation $\mathbf{w}[n] \in \mathrm{R}^{q}$ have the covariance matrices $\mathbf{Q} \in \mathrm{R}^{p\times p}$ and $\mathbf{C}[n] \in \mathrm{R}^{q\times q}$. The transition matrix $\mathbf{A} \in \mathrm{R}^{p\times p}$ describes the relation between the different time-steps. Finally, the matrix $\mathbf{H}[n] \in \mathrm{R}^{q\times p}$ is the so-called observation matrix. Using this model, the corresponding calculations for the KF are [Kay93]:

1. Prediction

$$\hat{\mathbf{s}}[n|n-1] = \mathbf{A}\hat{\mathbf{s}}[n-1|n-1] \in \mathrm{R}^{p}. \tag{4.3}$$

2. Minimum prediction MSE matrix

$$\mathbf{M}[n|n-1] = \mathbf{A}[n-1]\mathbf{M}[n-1|n-1]\mathbf{A}^{\mathrm{T}}[n-1] + \mathbf{Q}^{\mathrm{T}} \in \mathrm{R}^{p\times p}. \tag{4.4}$$

3. Kalman gain matrix

$$\mathbf{K}[n] = \mathbf{M}[n|n-1]\mathbf{H}^{\mathrm{T}}[n]\left(\mathbf{C}[n] + \mathbf{H}[n]\mathbf{M}[n|n-1]\mathbf{H}^{\mathrm{T}}[n]\right)^{-1} \in \mathrm{R}^{p\times q}. \tag{4.5}$$

4. Correction

$$\hat{\mathbf{s}}[n|n] = \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n]\left(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]\right) \in \mathrm{R}^{p}. \tag{4.6}$$

5. MMSE matrix

$$\mathbf{M}[n|n] = \left(\mathbf{I}_p - \mathbf{K}[n]\mathbf{H}[n]\right)\mathbf{M}[n|n-1] \in \mathrm{R}^{p\times p}. \tag{4.7}$$

Hence, after each time-step the current estimate of the state space variables is included in the vector $\hat{\mathbf{s}}[n|n]$. The MMSE matrix $\mathbf{M}[n|n]$ is composed of the theoretical MMSEs for each time-step. The KF is initialized with

$$\begin{aligned} \hat{\mathbf{s}}[-1|-1] &= \mathrm{E}\{\mathbf{s}[-1]\} = \boldsymbol{\mu}_s \\ \mathbf{M}[-1|-1] &= \mathbf{C}_s, \end{aligned} \tag{4.8}$$

based on a distribution of the initial state vector according to $N(\boldsymbol{\mu}_s, \mathbf{C}_s)$.

For the positioning application the parameters that have to be estimated are the UT position and its velocity in the 2-dimensional space, i.e.,

$$\mathbf{s}[n] = \left[x[n], y[n], v_x[n], v_y[n]\right]^{\mathrm{T}}. \tag{4.9}$$

Furthermore, we assume a mobility model that bases on random-walk behaviour with the corresponding transition matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4.10}$$

where T is the update period, i.e., one position is estimated each time-step. The state-space noise is

$$\mathbf{u}[n] = \left[u_x[n], u_y[n], v_x[n], v_y[n]\right]^{\mathrm{T}}, \tag{4.11}$$

The change of the UT position is controlled by process driving noise of certain variance. This affects also the velocity of the UT and can be set-up by the covariance matrix of the state-space noise vector given as

$$\mathbf{Q} = \begin{bmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_2}^2 \end{bmatrix}. \tag{4.12}$$

Currently, we do not consider any velocity information from the sources. However, also Doppler measurements could be easily included to obtain velocity information. Hence, currently the velocity states are assumed to be "hidden", i.e., for the delivered observations

$$\mathbf{x}[n] = \left[x[n], y[n]\right]^{\mathrm{T}} \tag{4.13}$$

we assume the observation matrix

$$\mathbf{H}[n] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{4.14}$$

Note that even though there is no velocity information delivered to the observation equation, the velocity is estimated implicitly by making use of the mobility model. The observation noise vector is given as

$$\mathbf{w}[n]=\left[w_x[n],w_y[n]\right]^{\mathrm{T}}, \tag{4.15}$$

where the corresponding measurement error covariance matrix $\mathbf{C}[n]$ has time-varying character due to the fact that the geometric constellations or number of visible sources may change over time, and thus, also the covariances of the errors can change.

4.3.4.2 Extended Kalman Filter

The main problem of the linear KF derived before is that it requires a linear state equation and a linear observation equation (and zero-mean Gaussian noise processes) to be optimum. Clearly, to track just the position of the terminal - based on recent position estimates and the mobility model - would result in such a linear relation. However, if we want to include measurements of all kinds (especially pseudoranges and TDOAs) that have a strongly non-linear character w.r.t. the current position, the standard linear KF is not a suitable approach to deal with this problem. For these situations an extended KF (EKF) is the better choice which is an extension of KF for non-linear problems. Generally, state space and observation equations can have a non-linear character. However, we just consider here a non-linear observation equation

$$\mathbf{x}[n]=\mathbf{h}(\mathbf{s}[n])+\mathbf{w}[n] \tag{4.16}$$

that includes the non-linear character of the observations or measurements. Hence, $\mathbf{h}(\mathbf{s}[n])\in \mathrm{R}^{q}$ is the non-linear dependency between the state space variables and the observations. Adapted for the positioning problem, this could be, e.g., the TDOA and/or pseudorange equations. In comparison to the linear KF, modifications of the standard vector KF equations are necessary. Therefore, we have to replace $\mathbf{h}(\mathbf{s}[n])$ in (4.16) by a linearization about the estimate $\hat{\mathbf{s}}[n|n-1]$ of $\mathbf{s}[n]$ in the current time-step, i.e.,

$$\mathbf{h}(\mathbf{s}[n])\approx\mathbf{h}(\hat{\mathbf{s}}[n|n-1])+\mathbf{H}[n](\mathbf{s}[n]-\hat{\mathbf{s}}[n|n-1]), \tag{4.17}$$

where the Jacobian observation matrix is calculated as

$$\mathbf{H}[n]=\left.\frac{\partial\mathbf{h}}{\partial\mathbf{s}[n]}\right|_{\mathbf{s}[n]=\hat{\mathbf{s}}[n|n-1]}, \tag{4.18}$$

i.e., it includes the derivation of the measurement model w.r.t. the variables of the state-space vector. Then, the linearized observation model can be rewritten as

$$\mathbf{x}[n]\approx\mathbf{h}(\hat{\mathbf{s}}[n|n-1])+\mathbf{H}[n](\mathbf{s}[n]-\hat{\mathbf{s}}[n|n-1])+\mathbf{w}[n]. \tag{4.19}$$

If we apply the standard KF equations to the linearized model in (4.19), the structure of the equations is the same as in for the KF, just the correction step (4) has to be updated to

$$\hat{\mathbf{s}}[n\,|\,n]=\hat{\mathbf{s}}[n\,|\,n-1]+\mathbf{K}[n]\big(\mathbf{x}[n]-\mathbf{h}(\hat{\mathbf{s}}[n\,|\,n-1])\big)\in \mathrm{R}^{p}. \tag{4.20}$$

Clearly, also the observation matrix has to be replaced by the Jacobian matrix. Note that compared to the vector KF the EKF has no longer the MMSE optimum behaviour as it is a property of the linear KF. Furthermore, Kalman gain and MMSE matrix can no longer be calculated "offline", i.e., apriori and without any knowledge of the observations. Hence, for the EKF these matrices have to be recalculated in every time-step depending on the current observations. In the EKF all available measurements (TDOAs and/or pseudoranges) are directly used as observed signals. This may be beneficial in environments where only a few sources are available. The assumed mobility model and the state space model are the same as described for the KF. As observables in the EKF, we use directly the TDOAs from the cellular unit and the pseudoranges from the GNSS unit if available. Hence, the observation vector is defined as

$$\mathbf{x}[n]=\left[d_{2,1}[n],d_{3,1}[n],\ldots,d_{N_{\mathrm{BS}},1}[n],\rho_1[n],\rho_2[n],\ldots,\rho_{N_{\mathrm{Sat}}}[n]\right]^{\mathrm{T}} \tag{4.21}$$

which results in an $(N_{\mathrm{BS}}+N_{\mathrm{Sat}}-1)$-dimensional vector, i.e., $N_{\mathrm{BS}}-1$ differential TDOA measurements performed with N_{BS} BSs are used, and N_{Sat} pseudoranges are available from GNSS measurements (cf. [WIND482]). The definition of the EKF is very general, thus, it is possible that the number of observables change "online", i.e., in each time-step. For instance, for different time-steps the number of satellites can vary due to, e.g., an occurring urban canyon situation. Therefore, besides the hybrid case also only cellular or only GNSS environments can be handled easily. Even the situation that no measurements are available for a certain time can be handled. In this case, just the prediction property of the filter is used. Where the observation vector $\mathbf{x}[n]$ is composed of the measurements, the vector $\mathbf{h}(\mathbf{s}[n])$ includes the according measurement models for the TDOAs and pseudoranges. Hence, it can directly be obtained from the measurement models defined in [WIND482], so the mapping function for the observable equation is

$$\mathbf{h}(\mathbf{s}[n]) = \begin{bmatrix} d_{2,1}(\mathbf{x})[n] \\ d_{3,1}(\mathbf{x})[n] \\ \vdots \\ d_{N_{\mathrm{BS}},1}(\mathbf{x})[n] \\ r_1(\mathbf{x})[n] \\ r_2(\mathbf{x})[n] \\ \vdots \\ r_{N_{\mathrm{Sat}}}(\mathbf{x})[n] \end{bmatrix}. \tag{4.22}$$

Correspondingly, the noise vector is defined as

$$\mathbf{w}[n] = \left[n_{2,1}[n], n_{3,1}[n], \ldots, n_{N_{\mathrm{BS}},1}[n], n_1[n], n_2[n], \ldots, n_{N_{\mathrm{Sat}}}[n]\right]^{\mathrm{T}}, \tag{4.23}$$

with the corresponding covariance matrix.

4.3.5 Simulation results for location determination

In [WIND482], link-level synchronization has been performed with at least three BSs at the same time and the navigation equation was then solved, i.e., the position information is extracted from the resulting TDOA measurement which is a non-linear estimation problem.

In Figure 4-43 the cumulative density function (CDF) of the estimated positioning errors is shown where the performance of the Minn algorithm is tested in a cellular network environment in the base coverage urban scenario under multipath conditions and the parameters given in [WIND6137]. The bandwidth is 50MHz, which results in a chip duration of $T_{\mathrm{Chip}} = 20\mathrm{ns}$. From a positioning point of view this chip duration yield an equivalent chip length of 6m. For the simulations we use the WINNER C2 wide area channel model. Furthermore, we assume a carrier-frequency offset of $v = 10$ normalized to the subcarrier spacing. Note that NLOS propagation is not investigated here. For positioning the strongest three BSs are used and an averaging is performed over 100 synchronization symbols which are equivalent to about 0.5s, i.e., every 0.5s new TDOA estimates are available. Additionally, we assume an SNR at the cell edge of -5dB where the cell radius is R=500m. The CDF in Figure 4-43 shows the probability that the root mean square error (RMSE) is below a fixed value err, where simulations of several UT positions and noise realizations were performed. We observe that for the static case and solution, e.g., in 90% of the cases the error is smaller than 95m and we see a significant gap between the coarse and fine estimation. The additional use of tracking is also investigated here. For that we have used the derived EKF tracking algorithm, where several UT tracks were generated and the EKF performance was analyzed using three BSs for positioning. The parameters for the mobility model which is implemented in the filters are given as follows: the covariance matrix of the state-space model is assumed to be

$$\mathbf{Q} = \begin{bmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_2}^2 \end{bmatrix} = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix} \tag{4.24}$$

Note that this yields tracks for pedestrian users, where the maximum velocity was set to 5km/h. Figure 4-43 shows the further performance improvement in terms of CDF for the dynamic case with EKF tracking. With this approach, we can achieve an error which is smaller than 75m in 90% of the cases.

Figure 4-44 shows the equivalent results for the microcellular deployment scenario [WIND6137] where the WINNER B1 channel model was applied. In this case, e.g., in 90% of the cases the error is smaller than 40m. The additional use of the EKF yields a further performance gain. Then, in around 90% of the cases the error is smaller than 25m. However, if NLOS propagation will be considered, performance degradation has to be expected. Here we have just considered the effects of multipath propagation and the detection of the weak signals.

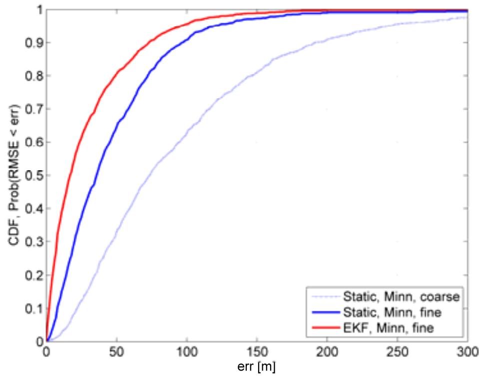

Figure 4-43 CDF for positioning, base coverage urban, Minn's algorithm for link-level synchronization, WINNER channel model C2

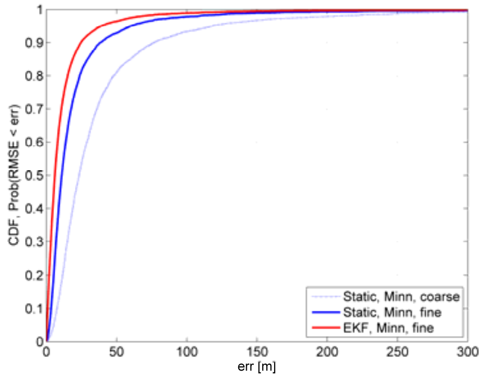

Figure 4-44 CDF for positioning, microcellular, Minn's algorithm for link-level synchronization, WINNER channel model B1

The inclusion of signals from GNSSs can further increase the positioning performance, especially in well-behaved situations (LOS access to several satellites). However, in urban canyons the performance can be limited [WIND482].

Figure 4-45 shows the performance of satellite based positioning in the static case and the dynamic case with EKF tracking. It is differentiated between GPS-only, Galileo-only, and combined GNSS (GPS+Galileo) for a fixed satellite constellation with standard error models described, e.g., in [WIND482]. We observe that the EKF can further improve the performance compared to static positioning. We achieve an accuracy which is better than around 7.5m in 90% of the cases for GNSS positioning using EKF.

In Figure 4-46 the positioning performance of GNSS in a critical situation is analyzed. This could be, e.g., an urban canyon scenario (cf. [WIND482]) where several satellites are blocked. For that, we assume that after 20 time-steps the urban canyon situation occurs for 100 time-steps, where only four, three, or two GNSS measurements are available for the EKF. We observe a high performance decrease if the number of visible satellites is below the critical four. However, even in these situations the EKF is suitable to track the UT for a certain time even though the overall position accuracy is low. In this context we refer to [WIND482] for an analysis of satellite visibility in urban canyon situations.

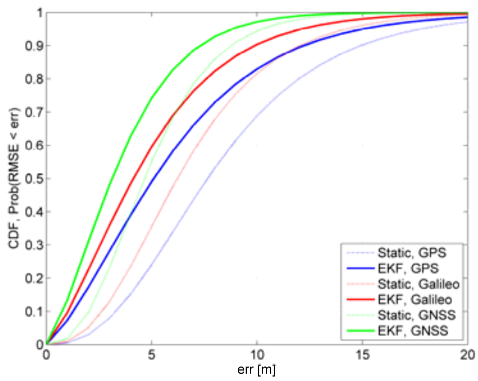

Figure 4-45 CDF for satellite based positioning, free space, static and dynamic case

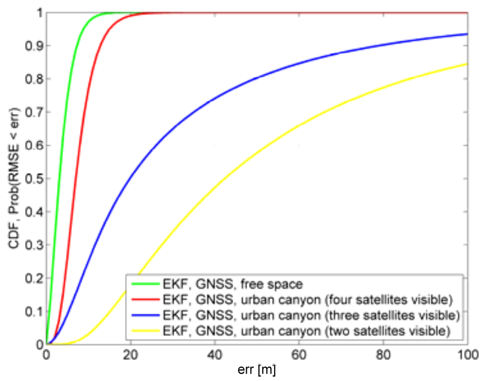

Figure 4-46 CDF for satellite based positioning, urban canyon situation with varying number of satellites

Finally, in Figure 4-47 the performance is plotted for a system that combines GNSS measurements with two TDOA measurements taken from simulations with the WINNER C2 channel model (with the parameters given above). All measurements are integrated in the EKF tracking algorithm. It can be seen that the additional performance gain is small when enough satellites are available. However, for only three or two satellites the performance can be increased and the lack of satellites can be compensated by the TDOA measurements. For instance, in 90% of the cases the error can be reduced from below 80m to below 60m if only three satellites are available, for two visible satellites the performance gain by additional TDOA measurements is even higher.

Figure 4-48 shows the corresponding results for TDOA measurements taken from simulations with the WINNER B1 channel model (with the parameters given above). We observe that, e.g., for only three visible satellites in 90% of the cases the error can be reduced from below 80m to around 25m. Hence, the missing satellites can partly be compensated by the TDOA measurements even though the error for the TDOA measurements is usually higher than the error of the pseudoranges for the GNSS measurements.

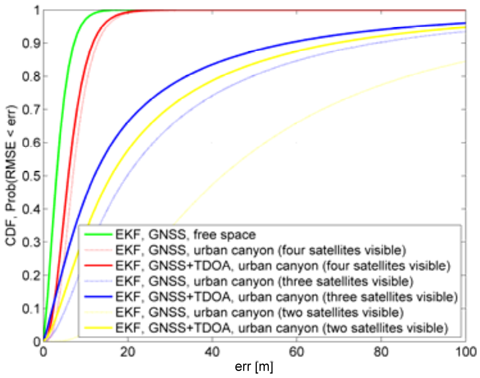

Figure 4-47 CDF for satellite based positioning, combination of GNSS with two TDOA measurements, WINNER channel model C2

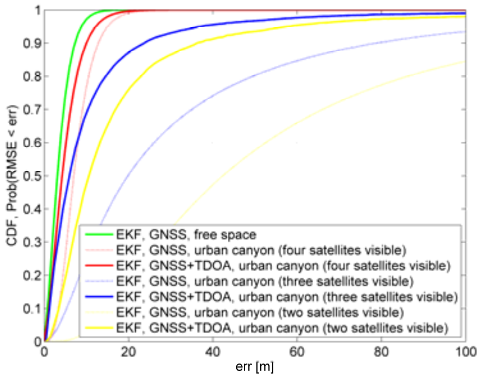

Figure 4-48 CDF for satellite based positioning, combination of GNSS with two TDOA measurements, WINNER channel model B1

4.3.6 Location based services

In addition to the classical and well-known user-side applications as like vehicle navigation, fraud detection, or automated billing, there are several system-side applications that can make use of location information [MTL+07].

One very important of such applications for the usage of location information are emergency calls. Common European agreements about accuracy requirements for location determination of emergency calls are not yet well defined. Nevertheless, the Coordination Group on Access to Location Information for Emergency Services (CGALIES) has developed a "Report on Implementation Issues Related to Access to Location Information by Emergency Services (E-112) in the European Union" [CGA02]. But currently there is no common agreement about these requirements. However, the United States FCC has stated accuracy requirements for E-911 emergency call services [FCC99] which were already discussed in [WIND482].

Location information can also be used to improve system operations of networks by including the spatial distribution of users and assets for communications (e.g., routing) and intersystem handover [WIN1D43]. An indispensable precondition to achieve integration of different networks is the possibility to allow for execution of handover between these systems. In a homogeneous system the scanning of other possible connections is triggered by the condition of the link, since an ongoing connection with good performance makes such a procedure dispensable. For an intersystem handover (ISHO), continuous surveillance is mandatory. Therefore the UT must scan all possible radio access technologies. The autonomous gathering of information by means of scanning may impact both the own and other transmissions. In all cases, the location-based ISHO in combination with the so-called hybrid information system (HIS) [WIN1D43] offers a great economic potential since participating devices can minimize or even avoid self-driven scanning procedures. The principle of the HIS presumes that each system collects data about the current link state within the covered cell and provides this information on request to UTs that are willing to change their connection within the same system or different systems. The basic idea behind the HIS approach is that each system reports about the current state, i.e., the link condition including, e.g., interference distribution. Together with a measurement report the location of the reporting UT within the covered cell is registered. The data is stored in a data base (DB) such that UTs of another system willing to change may request this information. Depending on the new target system and the current location of the UT, the UT is supplied with state reports of the same system type or another system, and subsequently may perform the handover, which is referred to as location-based ISHO since the location of the UT is exploited in the handover process. Obviously, the less accurate and precise the location information is the larger is the difference between the anticipated, i.e., retrieved measurement report and the real link condition in the target system after the handover. The achievable accuracy of stand-alone positioning is usually not sufficient, and hence, GNSS capabilities of a certain number of UTs are required for efficient ISHO processes. In the following (cf. Figure 4-49) this position-based, location-aided handover is explained in more details: Each active UT reports about the current link condition see (1). Together with the measurement report, the location of the reporting UT is stored in a DB. A UT that intends to perform a handover sends a request to its BS, see (3). The BS in turn acquires the corresponding measurement report from the DB, depending on the current location of the UT (4), and signals the handover decision (respectively related information that allow the UT to take the decision) to the UT (5). The UT can then perform the handover, which is marked by step (6) in Figure 4-49.

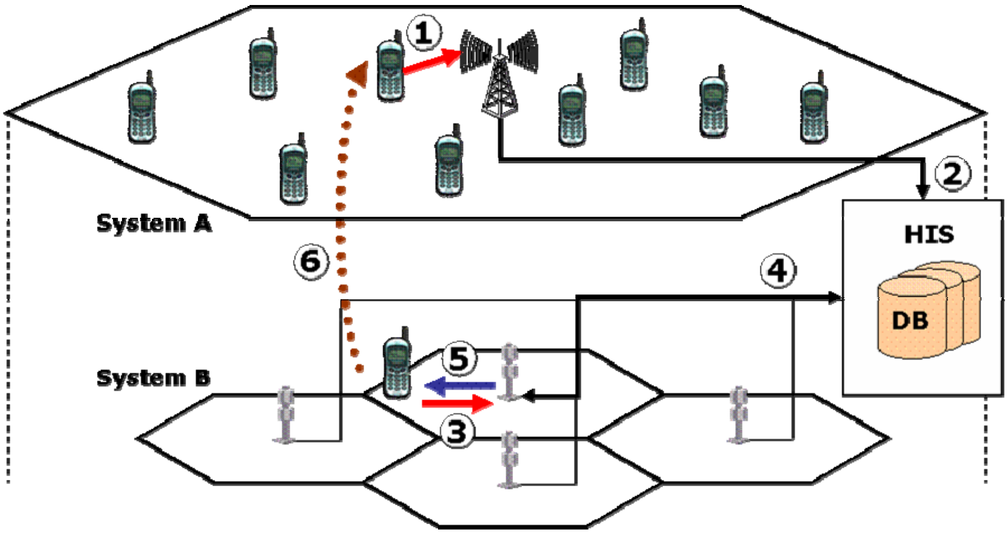

Figure 4-49 Location based ISHO using the HIS

Another system function that can exploit location information can be found in radio resource management (RRM) mechanisms to provide improved quality-of-service (QoS) to the user and save valuable radio resources for the network [HCD+02]. A network operator could not only use the location but also the speed and direction of the UT to further optimize the RRM, for the user admission, the user allocation to cells/networks and the load balancing between coupled networks. Location information can be used together with a known (from the operator) topology of the available networks in order to admit the user to the suitable network that has coverage in the location of the UT. This has an impact in the admission control procedure, where the network has to find the target system and/or cell for admission, so location information for the UT can be used to find the (suitable according to the UT requirements) systems for admission. Also, using location information together with speed and direction can help to admit the user in macro-micro cells. For example, fast moving users could be dispatched to systems with larger cells to reduce handover rate. A prediction of the future trajectory would enable to reserve the resources needed along the path of the UT. If within one system on the predicted path the required QoS can not be provided by one or several cells, the connection could be shifted to a system that can provide seamless QoS on the whole path in order to reduce inter-system handovers. The load balancing mechanism can also be optimized by location information, so the network can handover users to other systems that have coverage in an overloaded cell, by using the user's location.

4.3.7 Resume of location determination

In this section we have derived and analyzed the performance of an EKF based tracking algorithm for WINNER stand-alone location determination and the combination of WINNER based TDOA measurements with satellite based GNSS measurements. For WINNER stand-alone location determination with EKF the error is smaller than 75m in 90% of the cases for the wide area scenario (WINNER channel model C2) and below 40m for the microcellular deployment scenario (WINNER channel model B1). The simulations were performed under assumption that no additional NLOS bias affects the estimates, just the multipath bias was considered. If GNSS capabilities are available at the UT, the performance gain in terms of accuracy is very high, especially in situations where more than four satellites are available. Then, in a free space situation in 90% of the cases the error is below 12m for GPS-only and below 7.5m for combined GNSS (GPS+Galileo). However, in critical situations where only less than the required four satellites are available the performance loss is high. In these situations it is very beneficial to use additional TDOA measurements from the WINNER network. Simulation results have shown that if only three satellites are available the 90%-error can be reduced from 80m to below 60m if two additional TDOA measurements performed with BSs from a wide area scenario are used to compensate the lack of satellites. For the microcellular deployment scenario the 90%-error can be reduced from 80m to around 25m. Finally, emergency calls, inter-system handover, and radio resource management were identified and analyzed as system-side applications that can exploit available location information.

5. Cooperation Architecture

Radio Resource Management (RRM) aims to maximize resource utilization in the managed communication system. Optimization criteria are manifold and include for example maximization of spectral efficiency and achievement of user satisfaction comprising minimization of the number of dropped calls, provision of QoS demanded by users and optimization of (vertical) handovers – both handover delay and the overall number of handovers.

The tasks that RRM must usually perform in cellular systems are admission control, channel assignment, power control and handover control. These tasks have to be performed on basis of an estimation of the current system state. Inputs to such estimation are system dependent and may include for example the Received Signal Strength (RSS), the Signal-to-Interference and Noise Ratio (SINR), Distance to base stations, Transmit Power or mobile velocity.

In heterogeneous environments radio resource management does not only target the resource utilization of a single radio access technology but introduces cooperation between technologies and operation modes. Even more this cooperation will be of a decentralized nature as the long-term evolution of UMTS reallocates functionality from the Radio Network Controller (RNC) to eNodeBs (evolved NodeB).

For instance, current wireless local area networks use a Radio Resource Management (RRM) framework based on distributed architecture with simplified system functionality; because of low cost, self/easy deployment and high data rates are an important requirement, the Access Point usually has a direct connection to Internet. Conversely many current cellular systems use a centralized RRM approach to provide advanced system functionalities controlled by the operator: this architecture has been inherited from times when the only traffic component was voice, an application that don't need high data rates. Nowadays those wireless networks have to support an increasing demand of IP traffic and user data have to go through different nodes to reach the Internet (e.g. in UMTS through RNC, SGSN, GGSN).

Therefore, there is an increasing tendency in new cellular systems, as 3GPP LTE, WiMAx Mobile, IEEE 802.20 to increase system performance, namely higher throughput and lower latency, by reducing the number of system nodes, focusing on a decentralized architecture and also using "all IP" architectures to avoid trans-coding delays (e.g. in UMTS IP to ATM interfaces in the transport network).

3GPP is working on the so called Air Interface Evolution or Long Term Evolution (LTE) and also in the system architecture evolution (SAE). The network architecture is based on "all IP" networks. At Radio Access Level, only two elements are proposed: the base station, named eNodeB, and the gateway termed as aGW interfacing the core network. The eNodeB terminates most RRC/RRM protocols, management functionality and outer ARQ. There is a logical connection between eNodeBs, transmitting packet and context transfer for mobility management with minimal latency. LTE and SAE use radio handover for efficient and minimum packet losses.

5.1 Scalable Architecture

Figure 5-1 shows the cooperative radio resource management scheme. The primary goal is to introduce scalabilility and cooperation to the winner system. The cooperation allows interworking among different radio access networks and the efficient support of multi-mode and multi-system enabled devices. The proposed architecture is composed of a central coopRRM entity between WINNER and legacy systems and the optitional RRM server entities to support centralised radio resource management in the communication system.

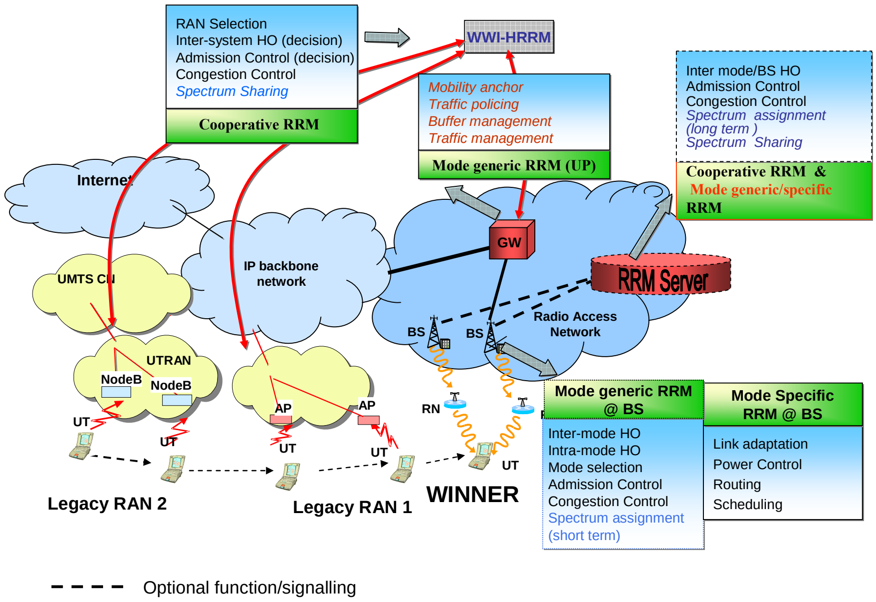

Figure 5-1: Functional allocation in logical entities architecture

5.1.1 Scalable Network Principles and Implications

WINNER system architecture holds the property of scalability that is a desirable property of the network, indicating its ability to either handle growing amounts of work in a graceful manner, or to be readily enlarged. For example, it can refer to the capability of a system to increase total throughput under an increased load when extra network entities (of the same types or of different types of the existing entities) that provide additional functions are added.

The scalable network implies the potential for function growth. As depicted in Figure 5-1, functions and interface between entities drawn in dotted lines are considered as optional functions. Without the existence of those functions, the system will work with the distributed mode. With the presence of those functions, the system will work under the centralised mode. Such scalability provides the freedom for the future operator to select the network being operated in a proper mode.

Performance of WINNER system will be improved after adding hardware, proportionally to the capacity added, is said to be a scalable system. An algorithm, design, networking protocol, program, or other system is said to scale if it is suitably efficient and practical when applied to large situations (e.g. a large input data set or large number of participating nodes in the case of a distributed system). If the design fails when the quantity increases then it does not scale.

5.1.1.1 Multiple solutions for Local Area BS deployment

Another aspect of scalability is the realization of the function allocation to physical node. According to WINNER solution on the logical node architecture [WIN2D428], the interface between the GW and the BS is specified as the logical interface.

Physically, it can be extended into multiple solutions. Figure 5-2 shows three typical solutions for the physical deployment of LA BS:

- Solution 1 gives the integrated version of the LA BS, where the logical GW function is integrated at the BS. Such solution is a typical one for Local Breakout. Terminals subscribed to the operators owning the LA BS are allowed to access to IP core network directly without going through an extra gateway.
- Solution 2 is a typical pico cell solution, whereas, the BS can be placed faraway from the GW. The interface between GW and BS is standardized, which allows the operator purchase network devices from different vendors.

- Solution 3 is another solution for local access. In that case, an LA BS will access the IP core through a GW (GW_G) which connects with the IP network and deals with policy enforcement, charging support and anchoring to non-WRAN system.

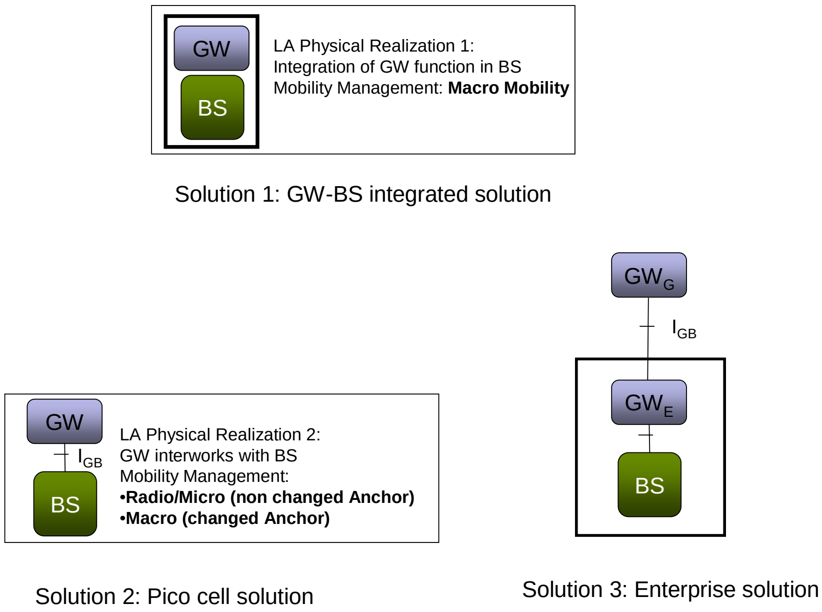

Figure 5-2: Alternative Solutions for locale area

The mobility management schemes will be adjusted accordingly to the deployment solutions. Mobiltiy management taken place under the assumption of solution 1 will be purely IP handover. The one under the assumption of solution 2 will be the mixture of radio handover and IP handover, as discussed in previous sections.

5.1.1.2 Benefits from optional network components

CAPEX will be saved when the operator choose network deployment without deploying the optional entities. The QoS experienced by the end user will not be degradated when the system does not require a center entity. On the other hand, in case the system is deployed in high tariff defined area, where the involving mobile users have high QoS demand, it will be beneficial for the network operator to deploy the central entity.

Such scalability designed by WINNER provides the feasibility for network migration. With the increasing number of users, the system can deploy additional optional network entity to improve the system performance.

5.1.2 Mandatory optional RRM server

Interface between server and the BSs are meshed connected. This allows saving cost for the deployment of the RRM server. The system can limit the controlling associations between the RRM server and the vicinity of the highly loaded BS, that can potentially benefit from the QoS improvement thanks to centralised RRM.

5.1.2.1 Distributed and centralized RRM, the RM server

The RRM server can give potential gains provided by the centralised joint radio resource control through the interfaces between the RRM server and the BSs. The system capacity gain obtained from the deployment of the RRM server is in principle the enlargement of the number of operational servers from the queuing model viewpoint, which therefore results in a higher trunking gain. On the other hand, by alternatively allocating the resources to call units among the interworking coexisting base stations, the load balancing effect among the radio networks is realized. In an interference limited network, such effect is very significant.

5.1.2.2 Potential gains provided by centralised RRM

The more servers available, the less the call blocking/dropping rate, meaning that there will be higher trunking gains thanks to the Admission Control over multiple available cells and modes. In addition, for a HCS (Hierarchy Cellular Structure) situation with overlapping e.g., WA BS and MA BS, the RRM server may allow traffic splitting over the involved BSs. It may result in an even less network response time due to the integration of radio resources [Luo05]. This gain can be termed as multiplexing gain.

5.1.2.3 Cost consideration for network solutions

However, deployment of RRM server can potentially increase the complexity and CAPEX. The WINNER architecture design is in favour of an optional RRM server in order to support scalable network solutions.

Strategically, the necessary functions are not allocated in an optional entity such as the RRM server. If the operator is able to deploy the RRM server potential performance gains can be obtained. This solution is applicable for hotspots and high end UTs concentrating service zones. Another option is to switch off the RRM server, when the direct interworking between the BS or the coordination by the GW still allow a running network.

5.2 Interworking of WINNER modes

5.2.1 A single GW pool controlling a numbr of deployment modes

WINNER system considers different modes being operated at the same service area interwork with each other at least under the coupling with the GW pool, as depicted in Figure 5-3. The inter mode interworking is based on two connection solutions, i.e., through the backbone network provided by the inter-connectivities among BSs and in addition over the air interface between overlapping BSs, that is however optional and on demand. This mechanim is explained in 5.2.2.

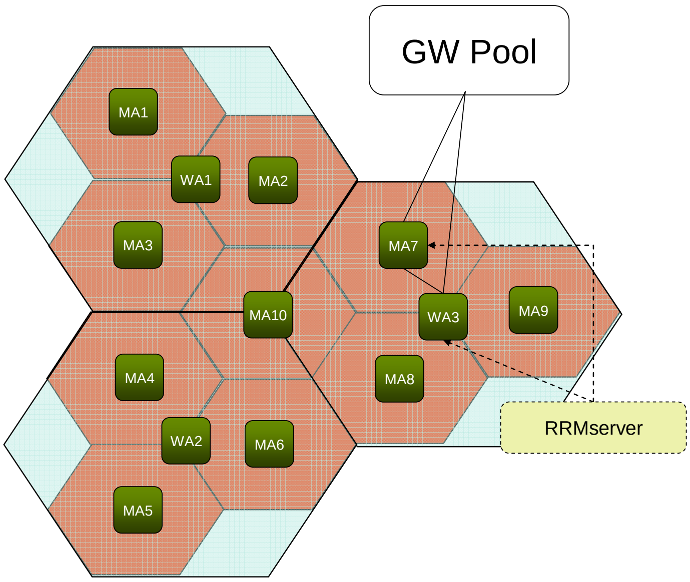

Figure 5-3: Single GW Pool controlling a number of deployment modes

5.2.2 Cooperation scenarios and interworking of widea area and local area BSs

Assuming that there are different types of BSs for the wide/metropolitan area and local area, we could restrict the extra functionality to the wide/metropolitan area BSs. Such network deployment is identical to the Hierarchy Cell Structure (HCS) architecture as classic mobile communication system. The difference is the over-the-air signalling between BSs.

In particular, the WINNER vision is that the cells of the different modes will coexist and overlap either completely or partially. This feature could be used in favour of the RRM architecture as the mode generic control plane functions that concern the coordination of the different modes/BSs could be moved to the WA BS and MA BS, making them responsible for the control and allocation of resources per wide area cell including all short range BSs (LA BS) that fall within its coverage. A requirement for such an approach would be the definition of a communication link between the WA BS and LA BS, this link could be either wired or wireless (e.g. part of the wide area mode interface).

Assuming this architecture there will be two hierarchical level of BSs, from one side the WA BS and MA BS and from another side their dependants LA BS.

The interworking between BSs uses the air interface carried by the wireless spectrum resource as the deployment of the WA BS and LA BS. The signalling link between WA BS and LA BS has been assigned previously.

5.3 WINNER – legacy systems interworking

One important goal for the WINNER project is to achieve global mobility management (we will consider handover management in particular) between the heterogeneous networks that compose WINNER and legacy systems. According to the different types of networks that are coupled and depending on the type of coupling, there are different integration schemes. In the following, we will consider WINNER cellular network as the default standard into which other possible networks may be integrated. In general, four different types of integration/coupling can be distinguished:

- Very tight coupling
- Tight coupling
- Loose coupling
- No coupling/Open coupling

5.3.1 Very tight coupling

In the very tight coupling case, WINNER and legacy systems are integrated at the RAN (Radio Access Network) layer, as depicted in Figure 5-4.

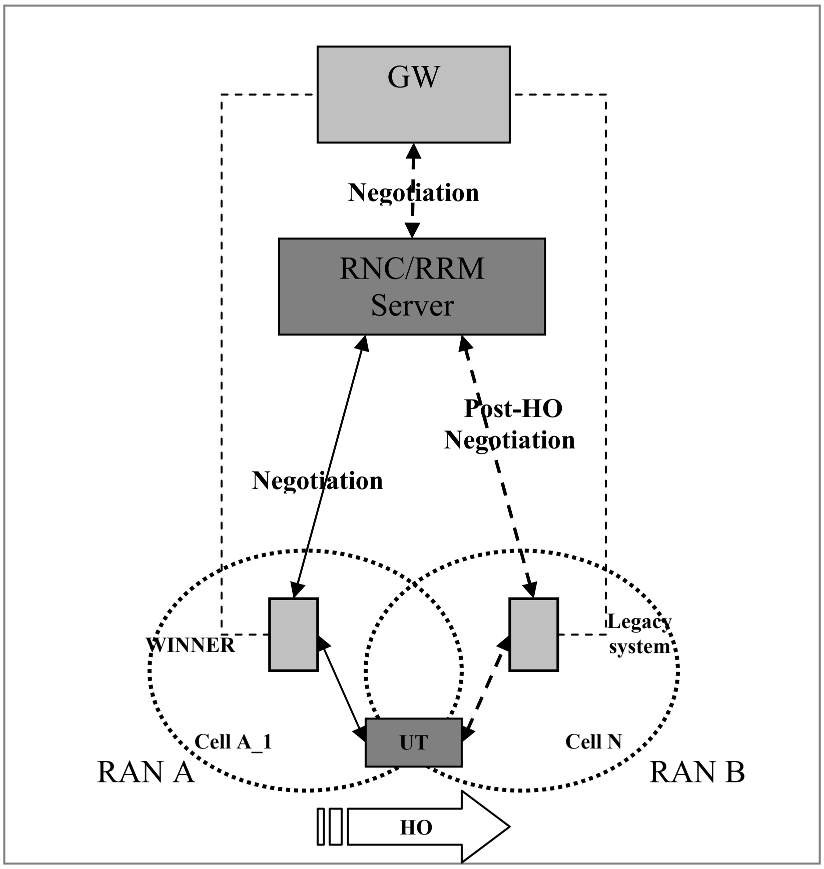

Figure 5-4: Very tight coupling scheme

The entities responsible for both radio access technologies are attached to the same RNC, multistandard BS or RRM server. Very tight integration is useful especially if the two networks are co-located and have overlapping areas of coverage. Resorting to such a coupling scenario is interesting where it is a matter of integrating local hot-spot networks without a great spatial extension, i.e. their area of coverage is comparable with the area for which the RNC-alike coupling point is responsible. Due to complexity and cost reason, this scenario is considered not to be the highest priority. Referring to section 4.1.6, when reconfiguration is embedded in the Multi-standard BS or in the RNC/RRM server, the network architecture is referred to as the very tight coupling.

Figure 5-4 escribes a very tight coupling architecture. We remark especially, the important role of RNC which represents the common integration point of the two networks. Thereby, the handover negotiations can be restricted to the RNC/RRM server and the UT. SGSN is involved in the handover management, only if some additional information not available in the central controller entity is needed.

To finish with, we can resume the advantages of very tight coupling scenario in the following:

- Possibility to establish a common resource management.
- This scheme is based on a limited number of entities required for vertical handover.
- Reducing delays to the minimum since all needed information are gathered within only one entity.

5.3.2 Tight Coupling

The tight coupling scheme integrates two networks at the RAN layer, as it is done in the very tight coupling scheme. The only difference is the use of two different RATs working together with a single core network. The coupling point is the GW for the packet switched (PS) domain, to which both RANs are attached. The GW in general includes the SGSN functionality supporting UMTS subsystem. This coupling scheme can be useful for co-located networks. Let's consider, for instance, the example of WINNNER and UMTS. In this case, both RANs are connected to the same GW/SGSN taking care of the connection management for both networks.
Even though this kind of coupling does not provide a common resource management, it is able to provide a fast and reliable inter-working for vertical handovers. This type of coupling is also covered by the function architecture illustrated in Section 4.1.6, when the coupling control function is allocated at the HRRM level.

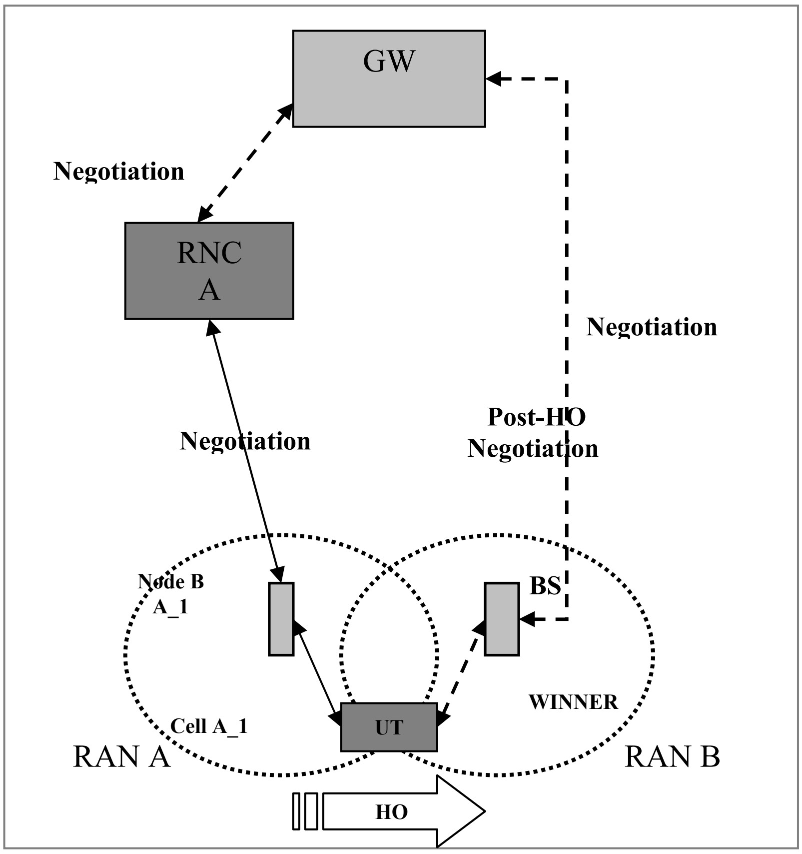

Figure 5-5: Tight coupling scheme

Figure 5-5 illustrates the different entities involved in a tight coupling scheme. We highlight the complexity of handover management comparing to the very tight coupling case. In effect, the serving and the target networks have different RRC controlling functions, namely the RNC for UMTS and BS for WINNER. Thereby; the UT must negotiate with both of them during the handover decision-making. The UT has just one contact to the serving RNC or BS which, therefore, consists in a relay point between the UT and the target RNC. Then, the negotiations or handover messages between the two RRC entities are done directly through the mobility anchor, i.e., the GW.

5.3.3 Loose Coupling

Unlike the two coupling schemes previously presented, the loose coupling scheme does not integrate the networks at RAN layer. Instead, the two different networks are connected via the core network, i.e. they are sharing the same core network infrastructure but using different GW or GW pools. One scenario is illustrated in Figure 5-6

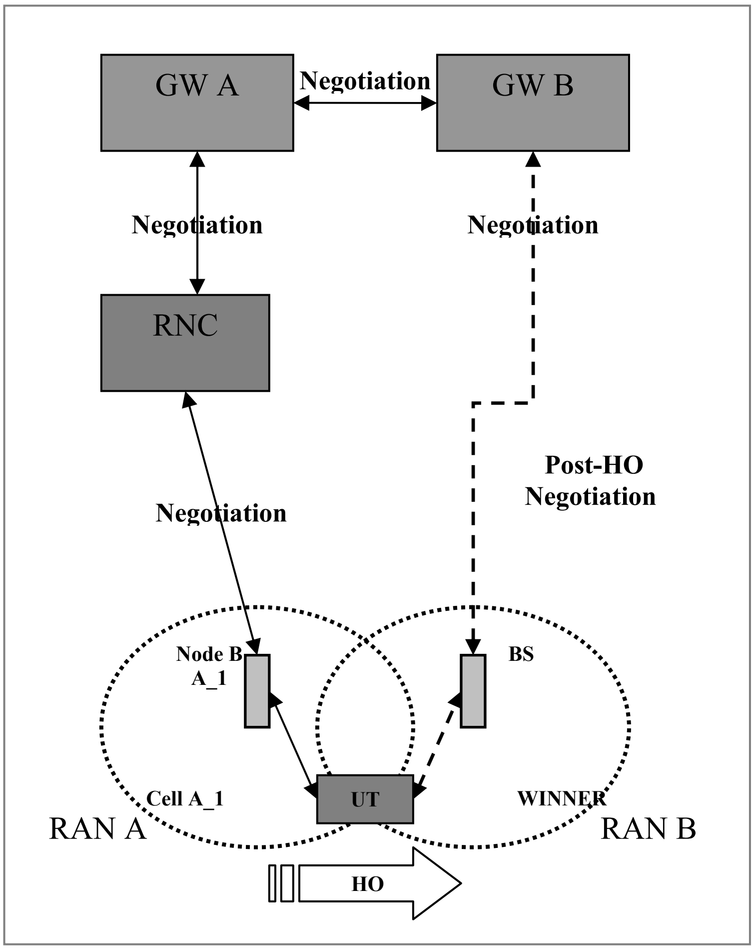

Figure 5-6: Loose coupling scheme

Loosely coupled networks offer a common interface for the exchange of information between the networks as required guaranteeing service continuity. Loosely coupled networks may be either geographically co-located or separated by a geographical boundary. The extent of information available for assessing the necessity and possibility of handover, however, is restricted. An exchange of information is possible, but requires considerably more time than in tighter integrated scenarios.

By consequence, handover negotiations are much more complex compared with the tight and very tight coupling scenarios. In effect, the UT starts negotiating with its serving RNC, which in its turn performs negotiations with the correspondent GW. The latter communicates with a GW of the second operator’s network or of another system. In the same way, the target network GW triggers corresponding RNC. As we can notice, the main difference compared with the intra-domain scenario is the participation of the GW, which is necessary while dealing with different networks.

5.4 Inter operator cooperation

WINNER architecture provides the interface (I_G) which defines the interwoking between WINNER and the legacy systems. The ARI (Ambient Radio Interface) interface refers to the terminology adopted within the project Ambient Networks. It is a sub-signalling interface of the I_G interface. The GW control plane has to support the necessary requests, measurement reports, and confirmations through the I_G interface.

An explicit CoopRRM server is located outside the WINNER RAN (see Figure 5-1). This entity can be mapped to a similar in functionalities entity within the project AN, namely the multi-radio resource management (MRRM) server. The inter-system interworking is enabled jointly by the CoopRRM and the I_G interface.

The generic RRM functions are allocated to the GW in order to facilitate inter-system interworking without causing too much signalling load.

5.4.1 Roaming and open coupling

Roaming as one of the existing inter-operator technologies can be considered as a candidate solution to solve the temporal shortage of the resource. The classical roaming, as provided by the radio network through inter-operator SLA (service level agreement) allows the mobile terminal free movement and reachability well beyond the coverage of single cells [Walke05]. Among the number of roaming types, i.e., across location area roaming, national roaming and international roaming, the roaming among overlapping PLMNs distinguished by operators provide connectivity for the mobile user even when their primary operator's network is fully loaded. From the Radio Resource Management (RRM) viewpoint, roaming is a feature where operators share the spectrum resources in their individual networks through the roaming mechanisms. Under the assumption of roaming, the networks between operators are coupled through the HSS/HLR (Home Subscriber Server/Home Location Register).

In comparison to that, a even simple simpler scheme called no-coupling or open coupling requires no integration of the networks at all. There is absolutely no mechanism for interchange of information between the networks. All actions have to be coordinated by the mobile station which acts as the only relay for the handover between the two networks. This option is possible for all types of networks; however, information about the new network and the possibility to perform a seamless and reliable handover is very restricted, unless a “make before break” strategy is employed for handover, which usually requires two radio transceivers.

5.4.2 Pools of GW and relationship to PLMNs

The pool of GW concept decouples the physical relation between a unique GW and a number of BS associated to that. Instead, each GW may associate to each BS in the pool area. Therefore, there is by default not necessary to employ inter GW HO within a pool area in case all GWs are eaqually balanced. Following that, the pool area is defined as an area in which the UT may roam without the need to change the GW. The GW capacity of a pool area can be scaled simply by adding more GWs. In contrast, in the classic hierarchical structure, each GW is associated with set of BSs serving their own location area, providing a direct mapping between a GW and the area covered by associated BS.

The operators can deploy multiple WINNER RANs e.g. in a larger country each covering a part of it (where a single network (e.g. a single RRM server) is not sufficient) and use regular macro mobility mechanisms using MIP protocol for handover between both networks. The MIP protocol can be used also for load balancing reason between GWs.

The inter-GW load balancing mechanism for the future network offers at least two-fold advantages. On one hand, the average flow processing time will be considerably reduced when the GWs are not equally loaded. On the other hand, the network can be designed in a more cost effective manner, which does not demand an over-engineered network using the conventional peak-hour based traffic analysis.
The network architecture jointly designed with handover offers an overall optimized performance. It enables the introduction of the hybrid handover mechanism which perfectly matches to the pool area deployment concept in the future network. The key reasons of using the pool area of gateways are the load balancing between gateways and avoiding unnecessary IP handover. It turns out that the maximum use of system resources and improvement of the system performance will be obtained.

5.4.3 Definition of Service Level Agreement

Service Level Agreement (SLA) is an official negotiated agreement between some numbers of partities, e.g., mobile users, network operators. The SLA can also be inter-operators. It is contracts the level of QoS being provided by one provider to its customer. It records the common understanding about services, priorities, responsibilities, guarantee, etc. with the main purpose to agree on the level of service. For example, it may specify the levels of availability, serviceability, performance, operation or other attributes of the service like billing and even penalties in the case of violation of the SLA.

In classic telecommunication system, SLA mainly refers to the contractual agreement between subscriber and the operator. Based on the previous contracted greements between these two parties, the operator should provide optimised QoS to the end user.

To support the proposed service optimization concepts, participating entities need to communicate and negotiate QoS parameters for their communication. This section puts focus on appropriate signalling within the IP Multimedia Subsystem (IMS) recently defined by 3GPP. The IMS utilizes the PS domain to transport multimedia signalling and bearer traffic [3GPP04]. The IMS enables operators to offer their subscribers multimedia services based on Internet applications, services and protocols. Transport services for multimedia signalling and user traffic are preformed by the PS domain. In terms of the ISO/OSI reference model, IMS is responsible for providing session layer functionality.

The session control protocol chosen by 3GPP for the IMS is the Session Initiation Protocol (SIP) defined by the IETF in RFC 3261 [Rosen02]. SIP was designed to support initiation, modification and termination of multimedia sessions. It is independent of the transported media. The IMS relies on other protocols (SDP, RTP, RSVP, RTCP) to describe the characteristics of transported multimedia streams (Session Description Protocol (SDP) [Hanley98]), transport streams (Real Time Protocol (RTP) [Schulz03]) and support QoS by resource reservation (ReSerVation setup Protocol (RSVP) [Braden97]) and evaluation of communication link characteristics during ongoing communication (Real-time Control Protocol (RTCP) [Schulz03]).

5.5 Inter Layer cooperation

5.5.1 Co-design of RLC layer, ARQ and HARQ

It has been established that the HARQ function alone may not be able to provide the required residual packet loss rates without uneconomic use of physical layer resources. For this reason a second level of retransmission should be taken as a protocol on top of the HARQ protocol. This turns out to be an RLC ARQ protocol added in the BS entity.

The location of the ARQ function cannot really be considered in isolation from other WINNER functions, for example in relationship to maximum number of HARQ according to maximum delay tolerance of the system. In WINNER system design, RLC ARQ is responsible for avoiding data loss during handover and segmentation of SDUs into PDUs. It also has implications that are outside of protocol performance such as the economic and implementation advantages, if any, of taking IP transfer directly to the Node B and avoiding investment in developing and deploying a central node capability, as the solution 1 stated in Section 5.1.1.

The PDU size on RLC layer is same as that on MAC layer PDU before the MAC multiplexing is performed. The impacts of channel-aware resource allocation on MAC HARQ and RLC ARQ, and interworking between HARQ and ARQ are of interests for investigations. Interactions among resource unit allocation scheme, HARQ/ARQ, and modulation-coding selection are of importance for system performance.

5.6 Inter-entity cooperation

WINNER network provides active inter-connections between relevant network entities. The inter-node connectivity provides interworking which potentially provides performance gains. Hereafter, some typical examples for inter-entity interworking are provided.

5.6.1 GW – BS

Inter GW-BS interworking considers both traffic behaviour and the system capacity. For example, the Admission Control functions (AC) include inter-system, inter-mode and inner mode Admission Control. They are allocated in BS and optionally in the RRMserver. Several scenarios can be described:

- AC in the BS that is mandatory for the radio admission control. In addition, WINNER supports the integrated coupling between BSs, which implies one BS directly controls another one. If WA/MA BS controls LA BS, AC is located in the WA BS.
- AC in the RRMserver for very tightly coupled inter-mode BSs (multi-mode) when the RRMserver is available. The term *very tight coupling* is another alternative of the coupling between BSs. In this case, the joint control between BSs is performed by the RRM server.

In order to support more intelligent AC, the congestion-avoidance based AC part is located in the GW for flow establishment and release according to the acceptable QoS level. This allows the GW to observe the potential congestion in the backhauling network before a radio admission control is carried out.

5.6.2 BS – Relay Node

Due to assumption of WINNER radio access network that supports operation at radio frequencies up to 5GHz. The radio propagation at such high frequency bands is more vulnerable to non-line-of-sight conditions and the achievable bit rates drop fast with distance. In order to provide cost efficient and flexible deployment solutions for different carrier frequencies and bandwidths, WINNER concept includes the option of using wireless relays. Relays can be applied to increase the coverage area of one BS, to increase the capacity of one BS, or to cover otherwise shadowed areas.

The interface between BS and RN normally takes place over a wireless link, but it could also be wired. This interface contains control traffic between the RN and BS and user plane information. For example intra-REC (Relay Enhanced Cell) routing and relay specific initial access signalling, advance forwarding of resource partitioning information to relays, and flow control signalling are supported by this interface.

Most of relaying support will take place in MAC. For example, supporting broadcast channel in all BSs and RNSs, and implementing resource partitioning are influenced by multi-hop deployment. It is also identified some key RLC layer protocols are also embedded relay, such as re-segmentation, RLC ARQ and status report.

5.6.3 Service server – RAN

Service server is expected to interact with the RAN through the GW functions. The control plane GW deals with functions as policy enhancement, packet detailed inspections and charging control. The user plane GW deals with IP layer convergence including the functions as header compression and ciphering mechanism. The WINNER GW communicates with service server to provide radio context dependent appropriate service according to user QoS demand (referring to SLA chapter).

5.7 Inter-Function Cooperation

All functions cooperate with each other to provide an optimized WINNER solution. For instance, functions such as Admission Control and Load Control are listed under the category of congestion avoidance control. To better explain the inter-relationships, Figure 5-7 and Figure 5-8 show the interworking among some of the RRM-related WINNER functions. Since congestion control itself is not an explicit function which evokes explicit WINNER air interface updates, it is classified to the network function and listed on the right hand side. The numbered functions are the priorities of execution.

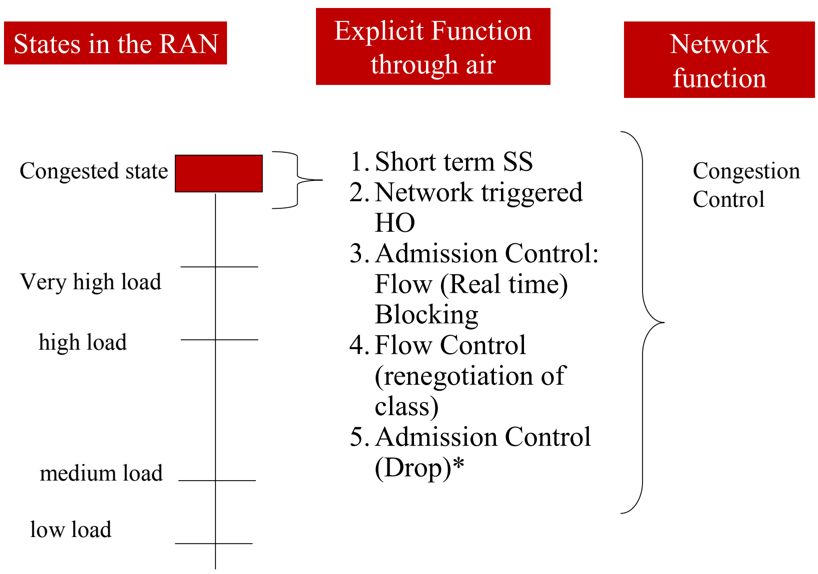

Figure 5-7: Interworking of selected inter-RRM function at congested State

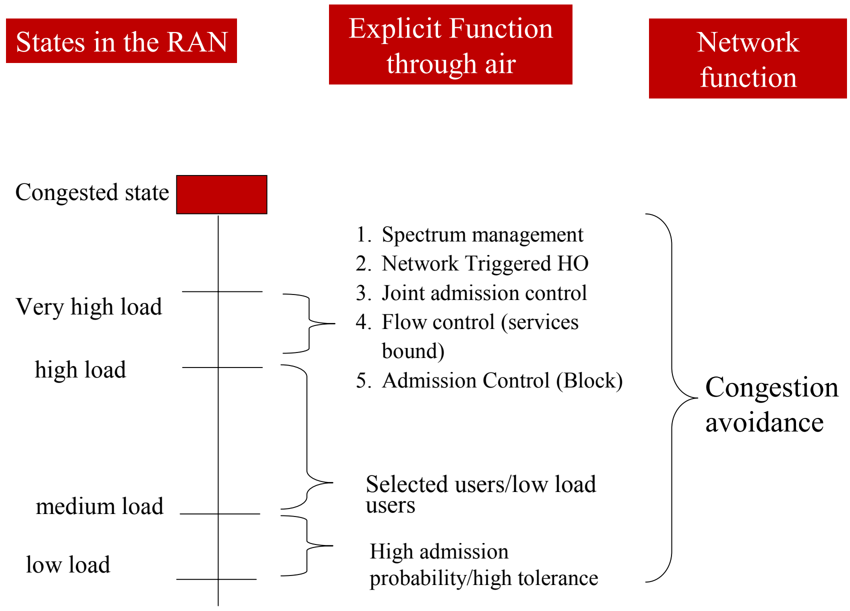

Figure 5-8: Interworking of selected inter-RRM function at high and low congested State

The term inter-mode handover and intra-mode handover are basically two types of intra-Winner system handover, where the modes are the WINNER deployment modes including wide area, metropolitan area and local area.
In the high load and medium load situation, the congestion avoidance function may trigger (network triggered) a handover by shifting some selected users to another cell/mode or frequency in order to avoid the congested state.
Under the scope of RRM, the functions need explicit RRC messages that are termed as 'Explicit function through the air' in Figure 5-7 and Figure 5-8. The rest are rather 'implicit' functions. The sequences of functions when a congested state or very high load situation occurs are numbered. The choice of execution priorities for some of the involving functions still need further study and therefore is marked by '*'. Figure 5-7 and Figure 5-8 show the function interworking congested state and states before congestion respectively.

6. Intramode, intermode and, intersystem, handover simulations

In this section we evaluate the performance of the WINNER handover mechanism. We consider both radio and IP handover, by evaluating intra-WINNER handovers (both intramode and intermode) as well as inter-system handovers (considering the case of an UT connecting to the legacy UMTS when losing the WINNER coverage).
Among the different possible WINNER RRM architectures, we focused on a distributed solution.
In the following, first we describe the proposed handover triggers, and then we evaluated the performance with the OPNET simulator. For a detailed description of the simulator please refer to [WIN2D482].

6.1 Handover mechanisms

6.1.1 Handover based on residual throughput and UT velocity

In this section we propose two different algorithms that should be jointly implemented to efficiently trigger a handover in the WINNER system. The first one, composed by two different triggers, can be used both for intramode and intermode handover, while the second one is specific for handover from local area to wide area.
The first algorithm we propose combines all the available parameters characterizing the quality of the communication (SINR, estimated instantaneous throughput and network load) and consists of two independent triggers, one for maintaining the wireless connection, and another for maximizing the user and network performance.
The first one is a wireless connection trigger which aims at guaranteeing an available wireless connection for the mobile station and only takes place when the actual connection degrades and is likely to be lost. This algorithm is reported in Figure 6-1. A SINR target is defined, to obtain a PER < 0.01 with packets of 1500 bytes on the more robust modulation scheme. By default the period of evaluation for the SINR is 0.2 s, but, when the SINR becomes lower than the target, it can be decreased to 0.02 s for an intensive evaluation. From the collected values an average SINR is determined. If the average SINR is lower than the target, the handover is triggered.

The network performance trigger instead combines the measured SINR and the network load in order to maximize the MAC layer performance. The metric used to activate the trigger is called "Residual throughput" and is defined as: Data Rate * (1 – PER) * (1 – Channel Occupation). The word "residual" means in this context, that if a part of network resource is already occupied by other users, then the handover decision is only based on the remaining bandwidth at the user disposal. The handover trigger is based on a comparison between the estimation of the residual throughput on the current cell (namely, Current_residual_throughput) with the one that could be achieved on another cell (namely, Target_residual_throughput).

If the ratio between Target_residual_throughput and Current_residual_throughput is bigger that a threshold,

Target_residuat_throughput / Current_residual_throughput > throughput_threshold

then the handover is triggered.

After extensive simulations we decided to set the *throughput_threshold* to 1.1, a value that avoids ping-pong effects and at the same time does not limit the gain that can be achieved with handovers.
Compared to the wireless connectivity trigger which periodically evaluates the link quality for maintaining the connection, evaluations of the network performance trigger are less frequent. Indeed, a time average of more samples can better express the network performance in "long term".
To derive the Data Rate, the Channel Occupation and the PER, the UT performs measurements in the used cells and on broadcast messages sent by the BSs of neighbouring cells. For a detailed description please refer to [WIN2D482].

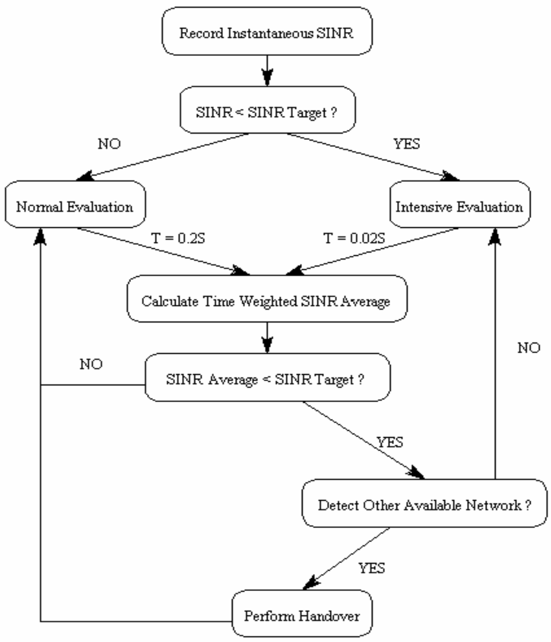

Figure 6-1: Wireless Connection trigger

The second algorithm we propose aims at avoiding frequent intramode handovers in case of high user mobility. A specific trigger for intermode handover has been introduced: if the velocity of the user terminal using a local area mode grows over a threshold Th_V, then a trigger for local-area to wide-area mode handover is activated. A good selection of Th_V, avoiding frequent interruptions of the communication and allowing exploiting the higher data rate of the local area mode, is 20 Km/h.

6.1.2 Handover based on fuzzy logic

For inter-system handover we propose to base UT decisions on fuzzy logic. We assume that, if no WINNER BS is available to guarantee wireless connectivity or good user performance, a handover to UMTS should be performed.

A Fuzzy Logic Controller is responsible for the inter-system handover trigger: it applies a set of predefined rules to the current system conditions. A neural network, instead, learns the FLC parameters from the resulting handover quality indicators. A representation of this scheme is presented in Figure 6-2.

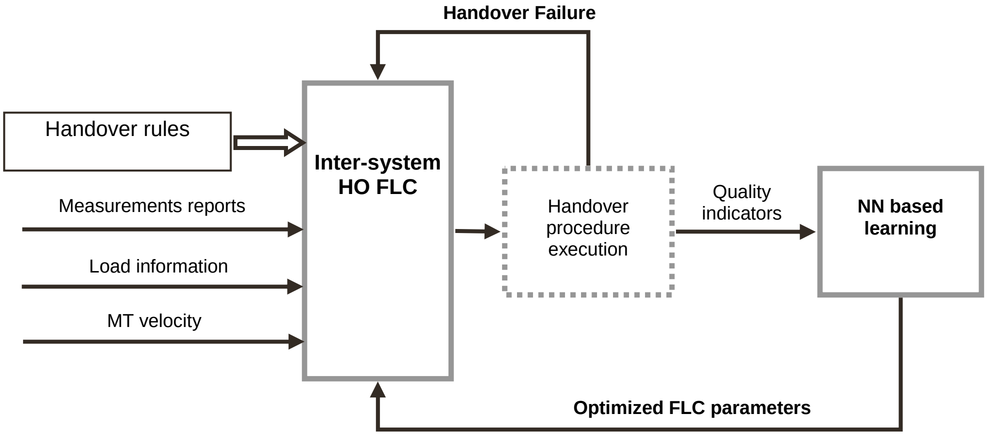

Figure 6-2: Fuzzy multi-criteria vertical handover scheme

A preliminary selection of target cells is performed before the vertical handover procedure. Signal level and load of different cells are filtered and then the cell with the best signal level is chosen. This target cell pre-selection reduces the FLC complexity and saves the processing time.

The considered handover criteria are based on different measurements: CPICH Ec/N0 (the received energy per chip divided by the power density in the band) for UMTS and the received signal power for WINNER, load information and UT velocity. The load criterion is very important in both handover directions. Load information on WINNER system avoids that users have low throughputs, since the increase of users' number results in the decrease of the bandwidth allocated to each user. Considering load information on UMTS instead prevents an UT from being downgraded or rejected by the Load Control mechanism. Velocity criterion is considered to avoid that a high-speed UT attempts to connect to a WINNER BSla.

We use membership functions composed of three fuzzy subsets representing the linguistic variables: low, medium and high. In the following part, levels 1, 2 and 3 refer, respectively, to low, medium and high levels. In short, if the signal strength on the current RAN is low and a target RAN with a signal level above level 2 and a load under level 2 exists, a handover is performed towards this RAN. If the load level on the current RAN is high and a target RAN with a signal level above or equal to level 2 and a load under or equal to level 2, a handover is performed towards this RAN. If conditions on the current system are optimal or if conditions on the best target system are worst, then the UT should stay on its current system.

Among these rules we included also the UT velocity: if the UT moves with a low velocity and conditions on WINNER LA are suitable then it should be handed over to WINNER LA.
We define a rule corresponding to the case where conditions on UMTS are still suitable and conditions on WINNER LA are optimal: in this case, a handover to WINNER LA is performed. This rule aims to enlighten UMTS load if a BSla is present. WINNER LA specific rules are based on velocity and could be summarized as follows: if the UT is on a BSla and it moves with a high velocity, it must be switched to UMTS if there is a target cell with suitable conditions.

6.2 Simulation results

6.2.1 Intramode handover

In this section we present results obtained with simulation showing the effectiveness of the proposed trigger for intramode handover. We first evaluate the delay of radio and IP handover; then, we show the gain in terms of throughput and delay achieved by efficiently triggering the handover, both in case of mobility and network congestion.

The fist scenario we refer to is represented in Figure6-3. Two BSs operating in the local area mode are available and one user terminal, initially connected to BS1, moves away at 5m/s thus entering the coverage of BS2. Different positions of the BSs have been considered: results reported hereafter are averaged over all the simulations performed.

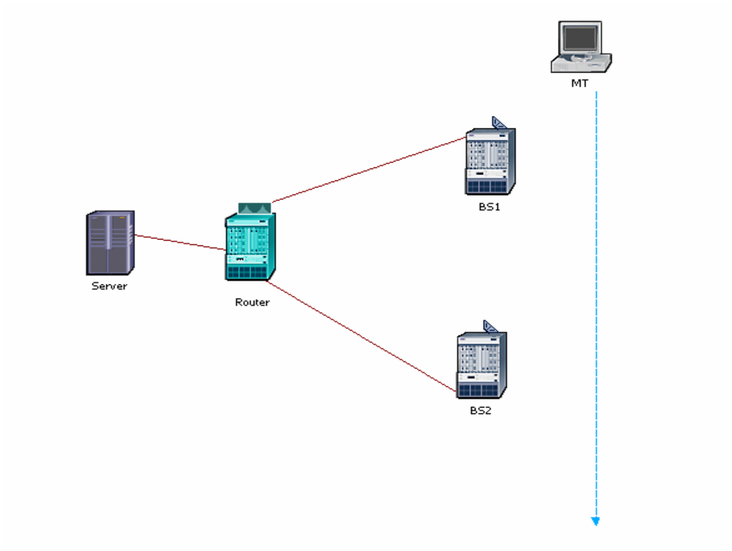

Figure6-3: Reference scenario for the evaluation of the trigger without cell congestion

In the first simulation set we considered long-lived ftp connections generated at the Server node.

We compare the proposed trigger to the trigger implemented in the legacy WLAN, activated at the disconnection from the current AP due to lack of connectivity. In the case of the legacy WLAN trigger, the search of an available BS is performed only when the connectivity toward BS1 is going to be lost. Conversely, when the WINNER trigger is implemented, the user terminal continuously performs measurements on the neighbouring cell. Without attending the loss of coverage: as soon as the target residual throughput increases the current residual throughput, the handover is performed.

6.2.1.1 Radio an IP intramode handover latency

Now, referring to the scenario presented in Figure6-3, we evaluate the handover latency of the WINNER system and we compare it with the latency of handover in WLANs. We considered two different deployment scenarios, with radio and IP handover respectively.

In the first case, the UT hands over between BSs controlled by the same gateway. In this case the IP address doesn't change and, as a consequence, this is a simple radio handover, which does not require the use of Mobile IP to maintain the connection.

In WLAN, the handover latency is due to the scanning procedure, the handover decision and the authentication/reassociation phase. Since our simulator does not implement the exchange of signalling messages, the values reported in the Table 6-1 do not include the authentication/reassociation delay. As reported in several work of measurements in WLANs, that delay usually varies between 5 and 15 ms, depending on the used card. To have a complete overview of the handover latency, these values have to be added to the simulation results. Shorter values are expected for the signalling time in WINNER, since scanning is performed on a distributed list of neighboring BSs before losing the connection and the authentication phase is not necessary.

Then we consider the case of an UT which hands over between BSs controlled by different gateways. In this case, an IP handover should be performed and the handover latency accounts also for the binding updates of the Mobile IP protocol.

The simulations results are presented in Table 6-1: we report the time between the handover initiation and the first data reception afer the handover. You can notice that also in case of radio handover, the latency of a WLAN handover is too high to offer seamless mobility. In the WINNER system instead it is equal to zero, since scanning and measurements are done rapidly and periodically on the cells of the neighbouring list only and no more at the handover time. We have to remember that the signalling delay has to be added to the values reported in the table.

	WINNER	WLAN
Layer2 HO	0	90 ms
Layer2+Layer3 HO	680 ms	780 ms

Table 6-1: Handover latencies

When Mobile IP has to be used to perform an IP handover, the latency dramatically increases, even if the WINNER case still offers better performance.

Note that, besides a reduced latency, the WINNER system offers further advantages: the introduction of the pool of gateways concept frequently lets to perform a radio rather than an IP handover, introducing a gain of hundreds of milliseconds.

6.2.1.2 Data rate and delay during intramode handover

We report now in Figure 6-4 the amount of received data as a function of the simulation time. We consider again the WINNER and the WLAN triggers. As you can see in the figure, the gain achieved with the proposed trigger is not only in terms of latency, but also of achieved throughput: the UT, exploiting information already collected, is able to select the BS that can provide better performance. Without attending the loss of coverage, as soon as the target residual throughput increases over the current residual throughput, the handover is performed. Thus, the handover to another cell allows the use of higher data rates, related to the higher SINR, since an adaptive Modulation and Coding Scheme (MCS) is adopted.

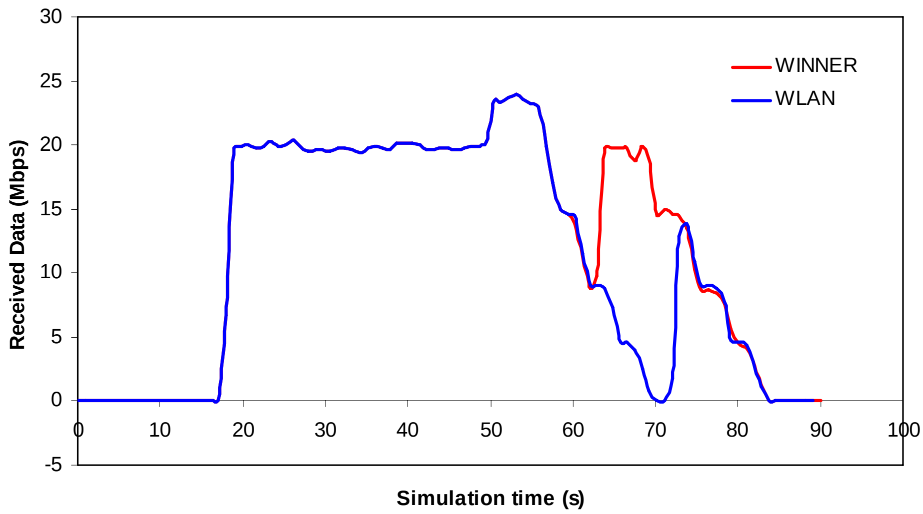

Figure 6-4: Received data as a function of the simulation time

Then we considered an interactive application like Voice over IP and we study how the two different triggers impact on the quality of the communication. We compare the performance derived by WINNER with that of legacy WLAN. Figure 6-5 reports the packet delay at the application layer as a function of the simulation time. At the beginning of the communication, which is at 13 s, the UT is in the coverage of BS1 and the delay of the voice packets is about 80 ms. From that moment on, the UT moves in directions of BS2. In the case of legacy WLAN, the handover to BS2 is triggered only when the coverage of BS1 is almost lost, that is at 68 s. The poor quality of the communication before the handover and the time necessary to scan all the channels to find the new BS introduce a high delay: you can observe the peak in the delay, that reaches 500 ms. Then the UT continues to move and exits from the coverage of BS2 at 85 s. There is no other available local area BS, BSla, so that at 85 s the communication is interrupted. This explains the new arise in the delay between 80 and 85 s.

The behaviour of the UT in the WINNER case is the same. The difference in the experienced delay depends only on the handover from BS1 to BS2. Indeed, with the WINNER trigger, the handover is performed immediately at 65 s, without degradation in the communication.

Figure 6-5 clearly shows that the WINNER handover avoids the peaks in the delay due to the interruption of the communication in the classical handover process.

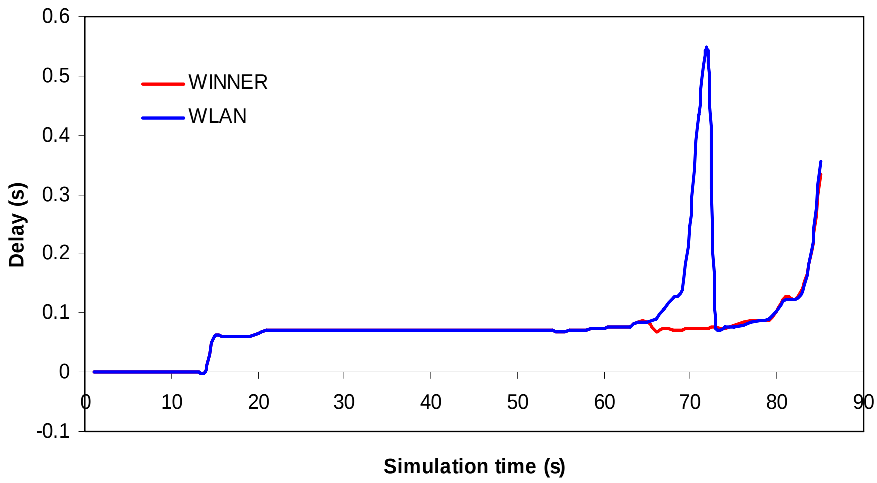

Figure 6-5: VoIP delay

6.2.1.3 Intramode handover triggered by residual throughput

Finally we show the performance gain that can be achieved performing handovers triggered by the residual throughput in different load conditions. Consider as an example the topology presented in Figure 6-6. The user terminal S0 is under the coverage of both BSla BS1 and BS2 and it is initially connected to BS1. Assume that, due to its placement, the SINR of the radio channel to BS1 perceived by S0 is 18.5dB. Then we vary the position of BS2 to achieve SINR in the second cell equal, lower and higher than that in the first cell. Moreover, we assume that the other nodes connected to BS1 generate a load of 5 Mbps, while the load of BS2 is varying.

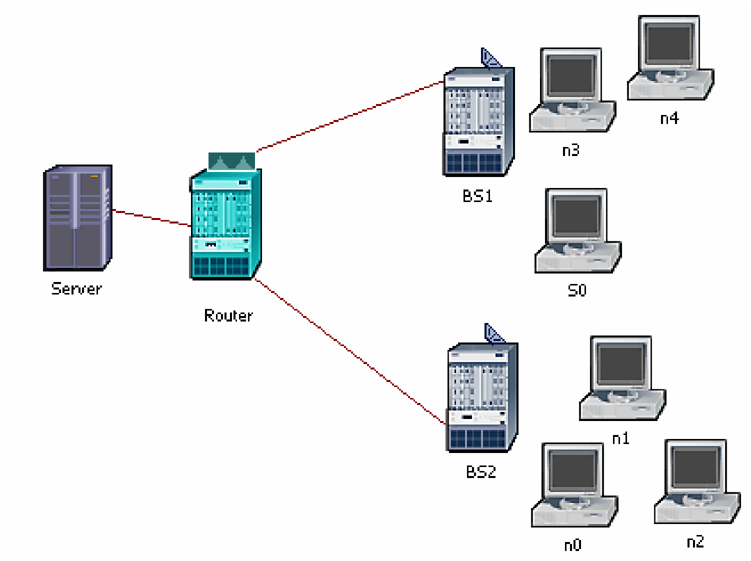

Figure 6-6: Reference scenario for the evaluation of the trigger with cell congestion

Figure 6-7 reports the layer 2 throughput achieved by station S0, as a function of the load ratio between the two cells, for different values of the SINR on the target channel (to BS2). The throughput achieved transmitting towards BS1, that is, if no handover is performed, is reported as a reference.
First of all, note that all the handovers result in a throughput gain. The estimation of the residual throughput indeed let the user correctly judge if advantages in performance can be achieved handing over. Moreover, we can gather that when the SINR on the target cell is higher than that one the current cell, the handover is performed until high values of load of the target cells. Otherwise, if there if no gain in the signal strength and thus in the used data rate, the handover is performed only when the load on the target cell is lower than that on the current cell.

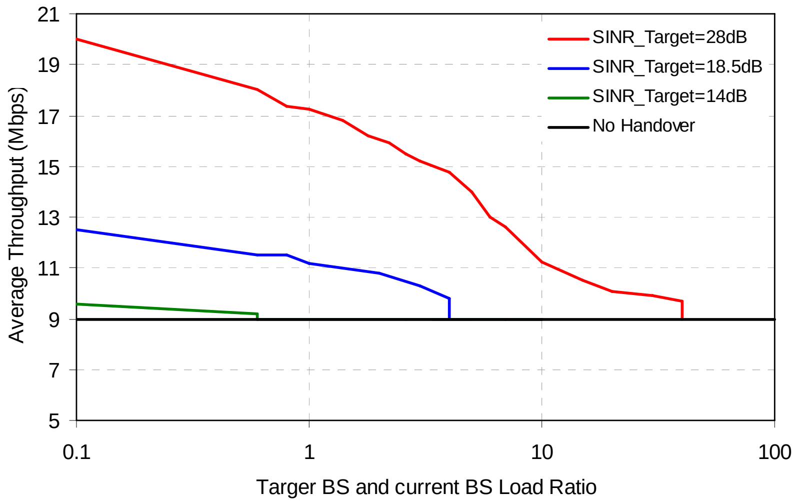

Figure 6-7: Average throughput of the user terminal S0 as a function of the load on the target channel for different SINR values

6.3 Intermode handover results

In this section we first evaluate the effect of the two different combined WINNER triggers (i.e., the wireless trigger and the throughput based on intermode handover) on UT mobility across the coverage of different modes. We consider the scenario represented in Figure 6-8: a mobile station crosses the wide area cell passing through two different local area cells. The mobile station activates the wireless interfaces in the following order: LA – WA – LA - WA.

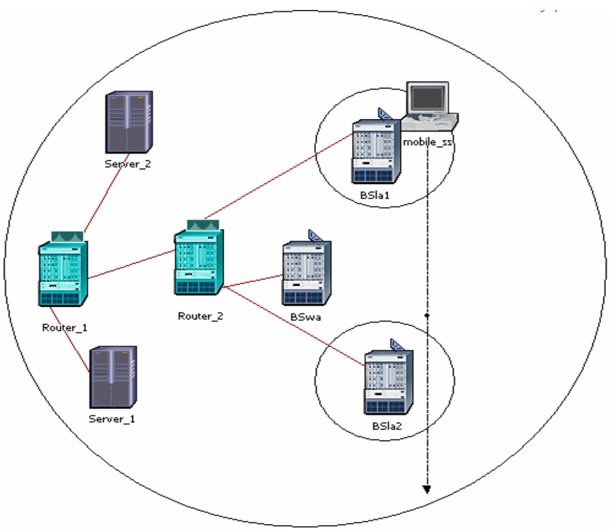

Figure 6-8: Handover scenario with residual throughput criterion

Figure 6-9 depicts the throughput achieved by the UT mobile_ss when using LA or WA modes. The UT, initially in the local area coverage, performs a handover to wide area when the throughput estimate on the wide area overcomes the one in the local area. Then, it enters the coverage of the second BSla and, when the LA throughput increases over the WA one, the second handover triggered by the network performance, from wide area to local area, is performed. Finally, the third handover from LA to WA is triggered by the wireless connectivity criterion. Indeed, the wide area throughput is lower than the local area throughput because the station is far away from the BSwa. For this reason the network performance trigger is not activated. In this case the wireless connectivity trigger lets to maintain the connection and avoids packets loss.

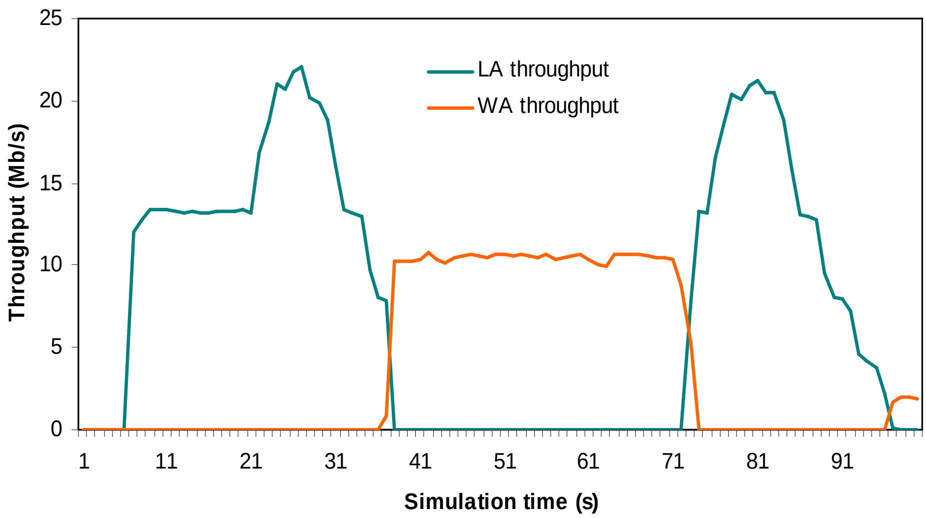

Figure 6-9: Estimated throughput of mobile_ss

We then analyze in Figure 6-10 the gain achieved implementing the specific intermode handover trigger based on UT velocity. We consider the following scenario: a road is covered by several LA BSs, located at a distance of 100 m one from the other and the user moves along this road that is totally covered by a Wide Area BS (BSwa).

Two cases are considered. In the first case, the specific speed-based trigger is not used: the UT, moving along the road, hands continuously over form one cell to the other since the wireless connection trigger is continuosly activated. In the second case, the specific speed-based trigger is implemented. Figure 6-10 reports the Cumulative Distribution Function (CDF) of the experienced Packet Error Rate (PER) for two different speeds, both above the threshold that activate the specific trigger. Note the amount of the PER reduction in case of use of the velocity trigger.

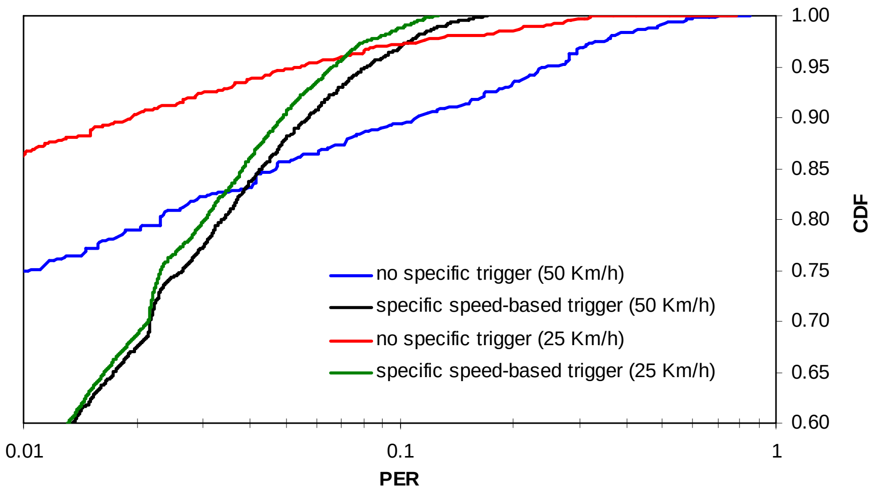

Figure 6-10: Cumulative Distribution Function of the Packet error rate at different speed

6.4 Inter-system handover

To evaluate the performance of inter-system handover algorithm, we consider three central and twelve neighbouring UMTS cells. Within each of the three central cells, there are three WINNER LA cells, as sketched in Figure 6-11. All UTs move at 3 km/h.

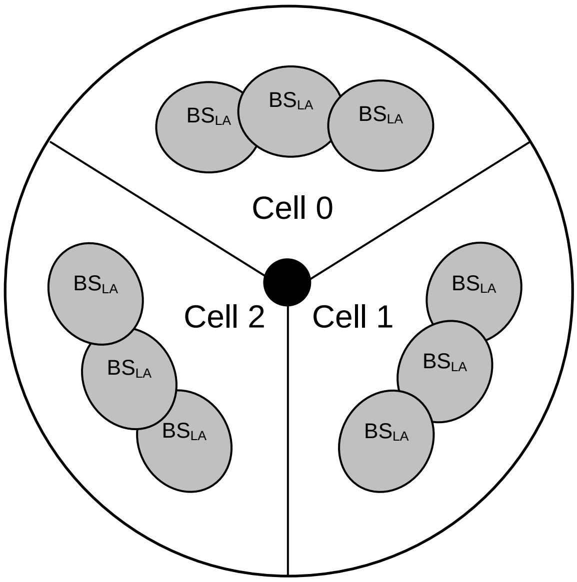

Figure 6-11: UMTS and WINNER LA simulation model

All users request a downlink Radio Access Bearer (RAB) of 384 Kbps and an uplink RAB of 128 Kbps. The maximum data rate that can be used to transmit to the BSla is 5.5 Mbps. The traffic model for each UT is a downlink heavy FTP service (100% get). The file size is 1.2 Mbits and the inter-request time is equal to 3s.
We compare the fuzzy adaptive inter-system handover algorithm to an algorithm based on fixed coverage and load thresholds. UMTS and WINNER LA have the same protocol stack from IP layer upwards. Therefore, we have chosen quality indicators of layers common to both RANs namely FTP download response time and TCP throughput.

The fuzzy logic based handover algorithm enhances the performances of the UTs. On Figure 6-12 and Figure 6-13 are depicted the histograms representing the FTP download response time and TCP throughput for 10 UTs. All histograms represent the distribution of the UTs over the horizontal axis intervals. The horizontal axis of the FTP download response time histogram represents the response time intervals in a logarithmic scale. The horizontal axis of throughput histogram represents the throughput intervals in a linear scale.

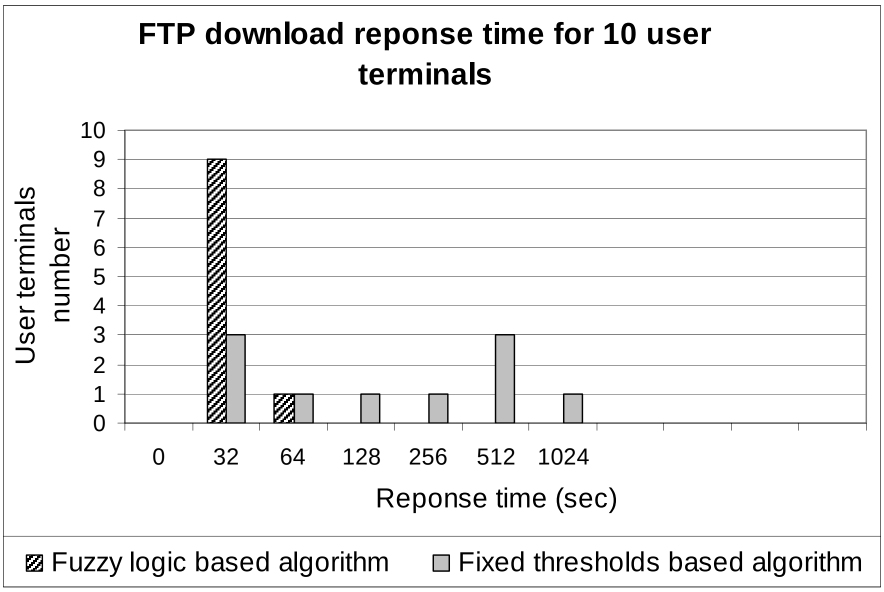

Figure 6-12: FTP Download Response Time histogram for 10 UTs.

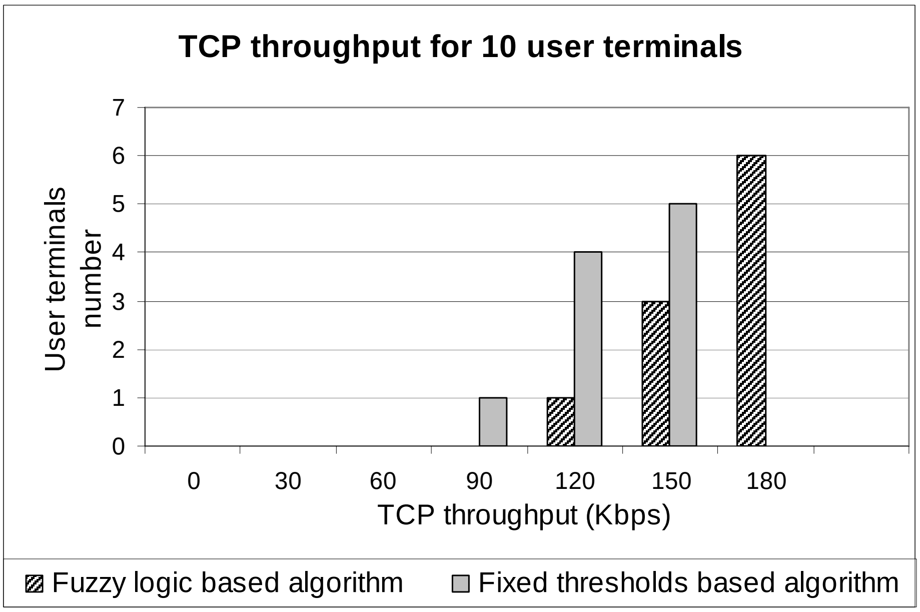

Figure 6-13: TCP throughput histogram for 10 UTs.

Finally we evaluate the performance of Radio and IP handovers. First, in Table 6-2 we report the duration of both radio and IP handover. As expected, radio handover delay results much smaller than IP handover delay.

Then, in Table 6-3 we present the E2E data delay experienced before and after a radio handover from UMTS to the WINNER LA mode. Before the handover, the UT is connected to the UMTS network and the downlink data packets flow through GGSN, SGSN and UTRAN. After the handover, the UT communicates with the WINNER BS and the downlink data packets are routed through the Gateway and the BS. This shorter path, together with the better performance of the WINNER LA mode, leads to smaller E2E data delay after the handover.
Finally, in Table 6-4 we present the E2E data delay experienced before and after an IP handover from UMTS to the WINNER LA mode. Note that also in this case the E2E delay after the handover results smaller than the delay experienced before the handover. However, compared with radio handover case, the E2E delay after the handover is higher: i.e., the tunneling of packets from the Mobile IP Home Agent to the Foreing Agent introduces a further delay.

	Radio handover	IP handover
Handover delays	180 ms	680 ms

Table 6-2: Handover Delay

	Before handover	After handover
E2E data delay	36 ms	21 ms

Table 6-3: E2E data delay in case of Radio Handover

	Before handover	After handover
E2E data delay	36 ms	27 ms

Table 6-4: E2E data delay in case of IP Handover

7. Conclusion

WINNER project has completed the studies on the definition of a next generation wireless network, resulting in a new set of system functions and also new radio network architecture, anticipating innovative and key technologies needed for the coming next generation of networks (e.g. interference mitigation, use the flexible use of the available bandwidth, etc..). The new radio network architecture allows new forms of network exploitation (e.g. sharing of BSs between operators).
In this document some of the features and innovations that correspond to the WINNER RAN higher layers have been described. In particular, the IPCL (IP Convergence Layer) and the RRC (Radio Resource Control) layer are of the main focus. The lower layers and some higher layer functions (as spectrum and interference mitigation) have been included in other documents of this project. Hereafter, a list of some innovative developments, described in this document is given:

- Scalable and hybrid RRM architecture provided by the RRM server.
- Fuzzy logic applied to intersystem handover.
- Coordinated radio and IP handover.
- Coordinated intersystem, inter-mode and intra-mode handover.
- Handover triggered by residual throughput and also on UT velocity.
- Admission Control based on backbone and radio load.
- Scalable context transfer for handover.
- Location determination based on the combination of WINNER radio and satellite signals.

The number of functions for the RRC layer was extended (e.g., the functions previously described, were mobility management, load and congestion control) respect previous deliverables, to include other essential set of system functions, necessary for the system operation. The list of new functions includes paging, cell selection and reselection, context transfer in idle and active mode, broadcast of system information, RRC connection establishment, flow admission control, ciphering and integration protection. The system functions are complemented by the so-called advanced functions, that offer additional features required for enhancing the performance of the basic functions even if the trade off is added complexity for a scenario of high density and degree of use deployments.

The overall protocol stack was described (including IPCL and RRC) in details and was mapped to the essential logical nodes (UT, BS, GW) considering a protocol split in the user and control planes. The needed information to support each RRC state (Disconnected, Idle and Active) was identified

The services, functions and protocols of the IPCL protocol layer have been described, namely, header compression and decompression of the different classes of Internet transport and network protocols (IP; UDP, TCP), transfer of user data between IPCL peer entities (typically the UT and the GW), the in sequence delivery of upper layer PDUs, and the duplicate detection of lower layer SDUs and ciphering of UP information.
The description of system functions and protocols associated to the RRC layer is the biggest part of this document. These have been divided in two groups, system concept functions and advanced functions.

The system concept functions (the majority of them belongs to the RRC layer), are those without which the system cannot work. The complete list of these functions is summarised here:

- Idle UT mobility management. The paging protocol supports UT mobility in idle mode but for this it needs some adaptation of the MAC superframe design to accommodate the paging indicator, as well as the protocol for cell selection and reselection, and context transfer in idle mode.
- Broadcast of system information. It has been identified the information to be broadcasted by the BCH channel to the UT in idle mode (e.g. cell ID, operator ID) and in active mode (e.g. spectrum sharing parameters, pointer to radio resources allocation table) was identified.
- Admission Control. An innovative approach to a two-level Admission Control was described. The motivation to adopt such an AC mechanism was that due to the high traffic capability of the radio interface, congestion could be located at the radio or at the backhaul, and the status of these two interfaces should be known, before an admission of new calls or sessions.
- Establishment, maintenance and release of an RRC connection. It has been described the first steps of network access, after the physical layer initial synchronization process as well as the interactions, after the establishment of the RRC connection, between the RRC and the NAS layer.
- Active mode micro-mobility. Mobility management in the active state is described by the intramode and inter-mode handover protocols.
- Intersystem handover. It has been presented a WWI functional architecture for the handover process during mobility management. The WWI architecture is a common approach that provides a complete

concept for a future generation communication systems (i.e., IST projects WINNER, Ambient Networks, E2R)
- Flow admission. Here, a description of flow admission interactions and protocols was given
- Load/Congestion Control. A description of flow control phases and its associated interactions (load monitoring, congestion resolution and congestion recovery phase) was given.
- Integrity protection and ciphering of RRC messages. For preventing the insertion and modification or RRC messages and also to keep the confidentiality of the contents. As the RRC layer now resides in the in the BS, it has been necessary to defined two different ciphering processes, one covering RRC messages (UT - BS) and other covering NAS and user plane (IPCL) messages (UT – GW / HSS)

The advanced functions (the majority of them belong to the RRC layer) that optionally could enhance system performance or provide new functions are the following:

- Hybrid and scalable network (described in detail in chapter 5) allowing traffic oriented network dimensioning. The planning of the network does not need the pure busy-hour based traffic pattern. Through on-demand centralised RRM introduction, the busy-hour traffic intensity can be dissolved by trunking gain given by resource sharing among collaborating network entities. On the other hand, due to inter connections among the GWs and BSs at the same service area, the network is easily migrateable.
- Distributed and centralized handover and hybrid handover (mix of the two types of handover)
- Radio and IP handover: Coordination of these two types of handover allows fast handover by default when two BSs belong to the same service pool without necessarily trigging IP handover. However the network owner may deploy the existing IP handover protocol without adding protocols to realise the load balancing among GWs that potentially reduce the network cost.
- Load control: in principle the load control can be deployed in many cases. On highlight in this document has proposed the load balancing between GWs at the same pool of gateways using the existing IP handover protocol as aforementioned.
- Policy based traffic and mobility management.
- RLC PDU context transfer during handover. Context transfer is needed to ensure seamless handover, i.e. without disturbing the ongoing session and also to make the overall process faster, i.e. no need to re-establish and re-authenticate the connection.
- Flow control between network entities, i.e., GW, BS and RN.

Policy management, Admission Control and RLC Context Transfer interactions:
- Since the GW helps the BS to perform the flow establishment and the flow management, the user may register in the network previously the most tolerant QoS for all possible QoS services as the preparation process for *Policy Enforcement*. The upcoming admission control and handover may be performed ccording to the tolerant QoS according to the acceptable Policy.

- Given by the policy, the two-stage admission control identifies the bottleneck of the network at the first step before admission control command is issued. This approach presented in this document solves the matching problem between Radio Access and Core network given by the Long Tail Distributed traffic characteristics of the future network. Alternatively to the algorithm in Chapter 4, the AC will be performed jointly by the RN, BS, and GW with the same priorities, similar as the IntServ procedure. Before a new flow is admitted, all entities in the e2e connection will be checked whether they can accept the new request. If this is the case, the UT will be informed respectively. This case however is only for new flows suitable, not for handover flows.

- In combination to the policy based mobility management, RLC context transfer between neighbouring BSs can be selected between PDU and SDU level. Even the RLC SDU level context transfer is selected, the high U-plane losslessness can be realized by intelligent inter entity flow control and local scheduling. We are able to conclude that by given suitable cross layer and cross function interworking, optimal systematic performance can be obtained.

For location service support, the WINNER stand-alone location information, TDOA timing information by OFDM synchronization algorithms is combined with timing information from positioning determination satellite systems (GPS, Galileo). Additionally, for further improvement, tracking algorithms for the solution of the navigation equation in the dynamic case were applied which improves the estimates in average. Furthermore, emergency calls, inter-system handover, and radio resource management were identified and analyzed as system-side applications that can exploit available location information. It has been identified the complementary location determination provides significant performance improvement.

The WINNER architecture has been continuously evolving since the earlier referenced versions (e.g., D4.8.1) towards increased adaptability to different deployment scenarios (Wide, Metropolitan and Local Area) and improved network performance and cost efficiency. The number of logical nodes has been reduced, reducing the signalling overhead and increase the user plane throughput. The challenge has been how to maintain all the control functionalities available in the previous RRM partially centralised structured architecture, and at the same time to offer high efficiency. It has been identified that the RRM functions must be jointly designed according to the network architecture. It has been also proven that with appropriate deployment of RRM functions, network deployment cost can be significantly reduced.

Finally, simulation results for the different types of handover: intramode, intermode and intersystem were presented, using algorithms based on the signal strength and cell data throughput, using a multimode/multi-RAN terminal and a common RRM entity that coordinate the different networks. It justifies that the RRC protocol designed for WINNER system introduces plausible performance catering for the future network and service requirements.

8. References

[3GPP04] 3GPP "IP Multimedia Subsytem (IMS) Stage 2", Technical Specification, TS 23.228 V5.13.0 (2004-12) (Release 5), 2004

[3GPP215] 3GPP TS 25.215, 3GPP; Technical Specification Group Radio Access Network; Physical layer - Measurements (FDD), Release 6, V6.4.0, 2005-09.

[3GP234] 3GPP TS 22.234, "Requirements on 3GPP system to WLAN interworking", Release 6, v6.3.0, 06/2005.

[3GPP300] 3GPP TS 36.300 V8.0.1, 3rd Generation Partnership Project;Technical Specification Group; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, Release 8, September 2007.

[3GPP322] 3GPP TS 25.322 V7.4.0, 3rd Generation Partnership Project;Technical Specification; Group Radio Access Network; Radio Link Control (RLC) protocol specification, (Release 7), September, 2007.

[3GPP821] 3GPP TR 33.821, "Rationale and track of security decisions in Long Term Evolved (LTE) RAN / 3GPP System Architecture Evolution (SAE)", Release 8, v0.3.0, 05/2007.

[Alcaraz06] Alcaraz, Juan J.; Cerdán, Fernando; García-Haro, Joan; "Optimizing TCP and RLC interaction in the UMTS radio access network", IEEE Network 20(2): 56-64, March/April 2006.

[Bajzik07] Bajzik, Lajos; Horvath, Peter; Korossy, Laszlo; Vulkan, Csaba; "Impact of Intra-LTE Handover with Forwarding on the User Connections", Mobile and Wireless Communications Summit, 2007. 16th IST, 1-5 July 2007.

[Braden97] Braden, R. "Resource ReSerVation Protocol – Version 1 Functional Specification", RFC 2205, September 1997

[Chen03] Chen, Yi-Chiun; Xu, Xiao; Xu, Hua; Gonen, Eren; Liu,Peijuan; "Simulation analysis of RLC for packet data services in UMTS systems", Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on, September 2003

[CGA02] Coordination Group on Access to Location Information for Emergency Services (CGALIES), "Final Report: Report on Implementation Issues Related to Access to Location Information by Emergency Services (E112) in the European Union", http://www.telematica.de/cgalies/, February 2002.

[FCC99] Federal Communications Commission, "FCC 99-245: Third Report and Order", http://www.fcc.gov/911/enhanced/, October 1999.

[GUO06] Q. Guo, X. Xu, J. Zue, H. Zhang, "A QoS-Guaranteed Cell Selection Strategy for Heterogeneous Cellular Systems", ETRI Journal, Volume 28, No. 1, February 2006

[Hanley98] Hanley, M. and Jacobson, V. "SDP: Session Description Protocol", RFC 2327, April 1998

[HCD+02] M. Hildebrand, G. Cristache, K. David, and F. Fechter, "Location-Based Radio Resource Management in Multi Standard Wireless Network Environments", Proceedings of the IST Mobile & Wireless Communications Summit, Thessaloniki, Greece, June 2002.

[Kay93] S. M. Kay, "Fundamentals of Statistical Signal Processing: Estimation Theory", Prentice Hall, 1993.

[KIM06] J.H. Kim, "On Delay-Bounded Multimedia Internet-Call Access Control in a WCDMA Common Channel", IEEE Transactions on Multimedia, Vol. 8, No. 5, October 2006.

[Lo05] Lo, A.; Heijenk, G.J.; Niemegeers, I.G.M.M.; "Performance evaluation of MPEG-4 video streaming over UMTS networks using an integrated tool environment", In: Proceedings 2005 International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS 2005, July 2005.

[Mikhovska07] A.Mihovska, B.Anggoro, S.Wardana, S.Kyriazakos, N.Prasad, "A Cognitive Approach to Network Monitoring in Heterogeneous Environments" WPMC 07, Jaipur, India, December 3-6, 2007.

[MTL+07] C. Mensing, E. Tragos, J. Luo, and E. Mino, "Location Determination Using In-Band Signaling for Mobility Management in Future Networks", Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Athens, Greece, September 2007.

[Navarro06] NAV06 [7] E.S. Navarro, V. W. S. Wong, "Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks", VTC Spring 2006, IEEE 63rd

[Protocols] http://www.protocols.com/papers/voip2.htm

[Racz07] Racz, Andras; Temesvary, Andras; Reider, Norbert; “Handover Performance in 3GPP Long Term Evolution (LTE) Systems”, Mobile and Wireless Communications Summit, 2007. 16th IST, 1-5 July 2007

[Rosen02] J. Rosenberg, et al. “SIP: Session Initiation Protocol”, RFC 3261, June 2002

[Sachs06] Sachs, J.; Khurana, B.S.; Mahonen, P.; “Evaluation of Handover Performance for TCP Traffic Based on Generic Link Layer Context Transfer”, Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on, September 2006

[Schulz03] Schulzrinne, H., Casner, S., Frederick, R. and Jacobson, V. “RTP: A Transport for Real Time Applications”, RFC 3550, July 2003

[Soldani03] D. Soldani and J. Laiho, “An Enhanced Radio Resource Management with Service and User Differentiation for UMTS Networks”, VTC 2003-Fall. 2003 IEEE 58th

[Song05] Q. Song, A. Jamalipour, “A Network Selection Mechanism for Next Generation Networks”, ICC 2005, IEEE

[WALKE05] B. Walke, Mobile Radio Networks, Networks, Protocols and Traffic Performance, 2005

[WIN1D43] WINNER I,” D4.3: Identification, Definition and Assessment of Cooperation Schemes between RANs“, Deliverable, IST-2003-507581, June 2005.

[WIND481] WINNER II, “D4.8.1: WINNER II Intramode and Intermode Cooperation Schemes Definition”, Deliverable, IST-4-027756 WINNER II, June 2006.

[WIND482] WINNER II, “D4.8.2: Cooperation Schemes Validation”, Deliverable, IST-4-027756 WINNER II, June 2007.

[WIND6121] Detailed WINNER II system trial plan for Phase III trials, Nov. 2007

[WIND6137] WINNER test scenarios and calibration case issue 2, Dec. 2006

[WIND6114] WINNER II, “D6.11.4: Final WINNER II System Requirements”, Deliverable, IST-4-027756 WINNER II, June 2007.

[Zhang04] ZHA04 [5] W. Zhang, “Handover Decision Using Fuzzy MADM in Heterogeneous Networks”, WCNC 2004, IEEE, March 2004