IST-4-027756 WINNER II

D4.8.1

WINNER II intramode and intermode cooperation schemes definition

Contractual Date of Delivery to the CEC:	**30/06/2006**
Actual Date of Delivery to the CEC:	**30/06/2006**
Author(s):	Emilio Mino[1], Albena Mihovska[2], Annika Wennström[3], Christian Mensing[4], Elias Tragos[5], Jijun Luo[6], Sana Horrich[7]
Participant(s):	*TID[1], AAU[2], KAU/CTH[3], DLR[4], NTUA[5], SM[6], FT[7]*
Workpackage:	*WP4*
Estimated person months:	30 PM
Security:	**Public**
Nature:	**Report**
Version:	1.0
Total number of pages:	**79**

Abstract:

This report identifies mechanisms and techniques for the efficient cooperation among WINNER and the legacies RANs and especially for the cooperation of WINNER modes namely; mobility management, congestion avoidance control and QoS based management. The applicability of fuzzy logic to mobility is also studied. The possible schemes for multi-hop ARQ are also analysed. To support these schemes an initial description of cooperation architecture is provided, later to be analyzed for obtaining the best candidate for the WINNER wireless network. And finally, the needed measurements to support the cooperation mechanisms are presented, with a section on measurements needed for location determination.

Keyword list: RRM cooperation mechanisms, RRM architecture, Triggers, Measurements, Mobility Management, Handover, Location determination, Admission Control, Congestion Control, Load Control, ARQ, Fuzzy Logic, QoS based management

Disclaimer:

Executive Summary

The main focus of this document is the initial definition of the algorithms of the cooperation mechanisms for joint management of WINNER radio resources between the three deployment scenarios (Wide Area, Metropolitan Area and Local Area). In particular, handover, admission control, congestion control and QoS based management. In a limited extend another scenario also have been considered, the cooperation of WINNER with the legacy RANs, based partially on the WINNER I results.

The WINNER system is a flexible system that will be able to handle the provision of future mobile multimedia services by providing wireless access from local area to wide area. It will offer an optimisation of capacity in the air interface by means of efficient RRM algorithms.

In order to ensure a high quality service provision to the user, a seamless handover between the WINNER modes and between WINNER and the legacy RANs is required. An initial study on the possible algorithms for the WINNER intermode and intramode handover between the wide area, metropolitan area and local area is being presented. Some advanced approaches as, the use of fuzzy logic to combine different inputs to the HO algorithms and obtain HO decision when several variables should be taken into account, and the use of Hybrid Information System, a specialized data base with UT measurements, are presented

Another point to highlight is the first definition of the cooperation architecture entities. The above-mentioned cooperation mechanisms between RANs will be located in a cooperative RRM (CoopRRM) entity for the cooperation with the legacy RANS. It is understood that each specific RRM related to a given RAN is located in a specific RRM (SRRM) entity that is connected to the CoopRRM. In the ACS, the generic and Intermode control functionality will be located; whereas in the BS, the intramode or mode specific functionality will be contained.

The ARQ protocols guarantee the reliability of multi-hop communication. Three ARQ proposals have been analyzed, namely layered ARQ, relay ARQ and multi-hop ARQ.

Measurements and triggers are an essential input of the RRM algorithms. The initial list of the measurements and triggers necessary for cooperation mechanisms is proposed, including the measurements for position determination and also the inputs for the WINNER cell load definitions

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
DLR Deutsches Zentrum für Luft- und Raumfahrt	Christian Mensing	Phone: +49 8153 282878 Fax: +49 8153 282875 e-mail: christian.mensing@dlr.de
FT R&D Division France Telecom R&D	Sana Horrich	Phone: +331 45 29 65 83 Fax: +331 45 29 41 94 e-mail: sana.horrich@francetelecom.com
KAU/CTH Karlstad University	Annika Wennström	Phone: Fax: e-mail: Annika.Wennstrom@kau.se
NTUA National University of Athens	Elias Tragos	Phone: +302107721511 Fax: +302107722534 e-mail: etragos@telecom.ntua.gr
SM Siemens	Jijun Luo	Phone: +49 89 636 45268 Fax: + 49 89 636 45591 e-mail: jesse.luo@siemens.com
TID Telefónica I+D	Emilio Mino	Phone: +34 91 3374428 Fax: + +34 91 3374212 e-mail: emino@tid.es

List of Acronyms and Abbreviations

3G	3rd Generation (Cellular System)
3GPP	3G Partnership Program
AC	Admission Control
AP	Access Point
ARQ	Automatic Repeat Request
BER	Bit Error Ratio
BLER	Block Error Rate
BS	Base Station
BSla	Base Station local area
BSma	Base Station metropolitan area
BSwa	Base Station wide area
DL	Down-Link area
FDD	Frequency Division Duplex
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HO	Handover
IEEE	Institute of Electrical and Electronic Engineers
IETF	Internet Engineering Task Force
IP	Internet Protocol
IAPP	Inter Access Point Protocol (802.11)
LA	Local Area (deployment concept)
MA	Metropolitan Area (deployment concept)
MAC	Medium Access Control
PER	Packet Error Rate
PHY	PHYsical layer
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RLC	Radio Link Control
RS	Resource Scheduller
RRC	Radio Resource Control
RSS	Received Signal Strength
RSSI	Received Signal Strength Indication
SGSN	Serving GPRS Support Node
SLC	Service Level Controller
SINR	Signal to Interference plus Noise Ratio
TCP	Transmission Control Protocol
TDD	Time Division Duplex
UL	Up-Link
UMTS	Universal Mobile Telecommunication System
UTRA	UMTS Terrestrial Radio Access
UTRAN	UMTS Terrestrial Radio Access Network
WA	Wide Area (deployment concept)
WLAN	Wireless Local Area Network

1. Introduction

This document presents WINNER RRM functionalities in the scope of WP4 / task8, including mobility support, congestion avoidance control and QoS based management. In WINNER I the focus on RRM was in the WINNER – legacy RANs functionalities and in WINNER II the focus is on the inter-mode and intramode functionalities. To support these functionalities, an appropriate inter-RAN RRM architecture was designed. The measurements needed for triggering the cooperation mechanisms (based in a so-called CoopRRM entity) have also been described. The descriptions have been based on the initial concepts as developed within the frames of the WINNER I. Here these concepts have been further extended in order to capture the overall advances in the design and typical characteristics of the WINNER air interface. The document presents some first ideas on how location determination can be applied to the WINNER terminal and several approaches for an efficient ARQ protocol in a multihop scenario.

This document is also aligned with the task 8 objectives as described in the Technical Annex:

- To refine the cooperation of WINNER II RAN with the legacy RANs, defined in WINNER I and to define the cooperation between the WINNER II modes defined also in WINNER I.
- Definition of the mode convergence protocol and architecture of the WINNER II modes. The aim is to facilitate the coexistence and cooperation of different modes in both logical and physical nodes of the WINNER II radio access network.

The document is organized as follows:

Chapter 2 describes the cooperation mechanisms amongst the WINNER modes and with the legacy RANs. First, the concept of WINNER modes has been presented. Then the cooperation mechanisms for the cooperation between the WINNER modes, these include mechanisms for, mobility management, and more specifically the management of handover, (a mechanism with special relevance to the project, because handover between modes and systems is an important prerequisite for the successful deployment of the WINNER system). Further, the cooperation mechanisms include mechanisms for congestion and avoidance control including admission and load control and QoS based management. These concepts have further been discussed on an intersystem level (cooperation between WINNER and the legacy RANs). A further step from the concepts presented in previous deliverables is the introduction of fuzzy logic methods for the performing of handover and QoS based management (other aspect as handover were studied in WINNER I)

Chapter 3 describes the cooperation in the user plane, specifically the ARQ protocols that guarantee the reliability of multi-hop communication. Three ARQ proposals have been analyzed, namely layered ARQ, relay ARQ and multi-hop ARQ.

Chapter 4 describes the measurements and triggers necessary for cooperation mechanisms, including the measurements for position determination and also the inputs for the WINNER cell load definitions (the cell load definitions for the legacies RANs were presented in phase I).

Chapter 5 describes the cooperation architecture. For the cooperation architecture the centralized approach was chosen as the most adequate for the cooperation between WINNER and the legacy RANs. Regarding the inter-mode and intramode the options of centralized and partially distributed architectures are discussed, presenting the trade-off between efficiency and complexity.

2. Cooperation mechanisms

2.1 WINNER modes concept

The WINNER radio interface is developed to adapt to a wide range of situations and environments, e.g. ranges, mobility, user densities, and datarates needs. The adaptation of the RAT might require different parameterizations or use of different algorithms. Certain combinations of parameter or algorithm assignations or ranges of parameter or algorithm assignations may be referred to as Modes. Modes are used in the project as a synonym for adaptivity of the system to different application scenarios, radio environments, spectrum bands, etc. Related to RRM architecture, the different deployment scenarios defined will be the basis for WINNER RRM modes, here we investigated the RRM functions (mobility management “HO”, admission control, QoS based management, Load/congestion control, Spectrum Control/assignment, Mode/RAN selection) among the different scenarios, i.e. Wide area, Metropolitan area and Local area. Figure 2-1 illustrates the WINNER RRM modes.

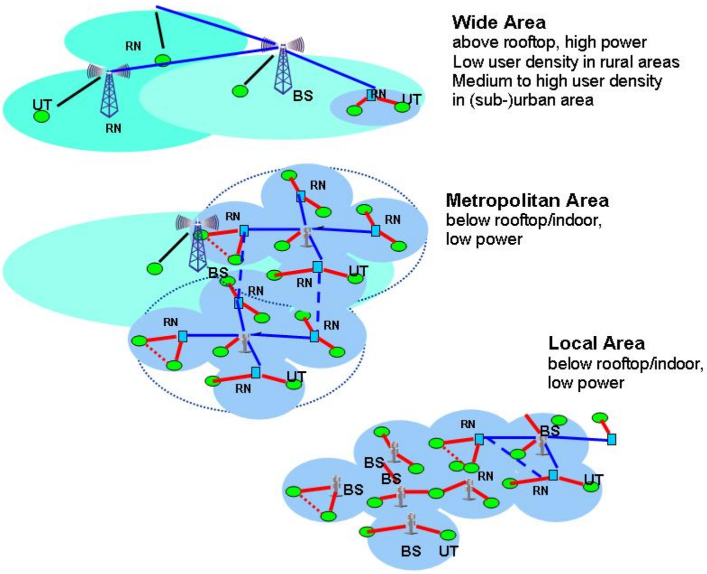

Figure 2-1: WINNER deployment Scenarios (WINNER RRM Modes)

WINNER uses two physical layer modes, i.e. FDD and TDD, for the three deployment scenarios: Wide Area (WA), Metropolitan Area (MA) and, Local Area (LA).

Table 2-1 presents the relationship between the different Base Stations (BSwa, BSma and BSla) that serve each deployment scenario, the type of physical layer mode, the type of spectrum used, mobility characteristics and the data rate offered by the deployment cell.

Types of Base Station	PHY layer mode used	Spectrum	Mobility support	Data rate	Cell size
Wide Area (BSwa)	FDD	Licensed	High <350 km/h	Medium (FDD)	High
Metropolitan Area (BSma)	FDD and TDD	Licensed	Medium <70 km/h	Medium (FDD) or Highest (TDD)	Medium
Local Area (BSla)	TDD	Licensed and unlicensed	Low < 5km/h	Highest (TDD)	Low

Table 2-1: Types of Base Station considering the deployment scenarios

One conclusion of this table is that the type of the physical layer mode used by one BS doesn’t identify the type of deployment scenario it is serving, for example an FDD BS can be deployed for a MA or WA cell. In the case of handover, and other cooperation mechanisms we should rely on the type of deployment of the base station not on the type of PHY layer mode used by the BS.

Regarding the mobility support is clear that the medium to high speed users (>70 km/h) can only be served by a WA deployment. Therefore the user speed is a parameter that can force UT handover. The future WINNER system may deploy a Hybrid Information System (HIS) to identify and restore user profile, where the mobility parameters such as velocity and mobility as well the relative location w.r.t. the network deployment are included. In Section 2.2.1.7, advanced location based HO is explained.
It is expected that the most demanding and futuristic services can be supported only by LA deployment, but an important factor to be taken into account is that the maximum datarate offered to one user depends on the cell congestion (number of free chunks), and therefore sometimes the maximum datarate could be offered by a MA or WA deployment.
The highest spectral efficiency will be offered in this order, LA MA and WA, meaning that with similar radio link conditions and cell occupation the UT will be better served also in this order.
Assuming different types of BSs for the wide/metropolitan area and short range mode, we could restrict the extra functionality to the wide/metropolitan area 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 BSwa and Bsma, making them responsible for the control and allocation of resources per wide area cell including all short range BSs (BSla) that fall within its coverage.

2.2 Cooperation amongst WINNER modes

The WINNER system has been designed for different deployment scenarios (wide area, metropolitan area and local area) in order to cover all the users' demands and serve them with the best possible quality of service. WINNER should be adaptive to all kinds of application scenarios, radio environments and spectrum bands and for those reasons, there have been defined different "modes" of the WINNER system. This adaptation of the system might require different parameterizations or use of different algorithms. As defined in WINNER phase 1, the modes of the WINNER system are certain combinations of parameter or algorithm assignations or ranges of parameter or algorithm assignations.

Since the WINNER system is supposed to be one unique system and not several separate systems each one for each mode, the modes should be able to cooperate with each other. In order to allow the cooperation among modes there are many Radio Resource Management mechanisms are needed, that will actualize and support the cooperation. In the next sections we present a first description of the RRM mechanisms that will ensure the cooperation among the WINNER modes. These mechanisms include Mobility Management (intra-mode and inter-mode Handover), Congestion Avoidance control and Quality of Service Management.

2.2.1 Mobility management

Mobility management techniques support the user mobility, including the traffic balancing which is essential for a network to use efficiently the resources of the system. In WINNER system, where different modes will be operating together, the traffic balancing between modes will be a very important mechanism to keep the network operating in a normal state serving the users with highest possible QoS. The intra-mode and inter-mode handover will be key mechanisms to support the traffic balancing inside the WINNER system. The most essential trigger for handover is the signal strength (like the handovers in the currently deployed networks), but also on load and service based information, mobility (velocity), user's subscription profile, user's preferences, terminal capabilities, user's environment (indoor, outdoor, etc.), location, interference, and statistical information. Since the different WINNER modes are defined in order to exploit the different deployment scenarios, the mobility and the location of the users will play a very important role in the handover process.

Some of the concepts for cooperation between WINNER and the legacies RANs developed in phase I can be extended the cooperation amongst WINNER modes as the handover scenarios and the generic handover flow in which the UT and BS will periodically perform measurements/calculations or exchange information regarding the current or neighbouring cells of the same or different RANs. Also, the higher layer triggers are expected to be activated either by AP calculations on the cell status or by information sent by the monitoring entities.
Figure 2-2 present this concept, in which the periodic measurements can trigger a handover process, but also a higher layer trigger (e.g. cell load), then the UT request to the network elements information on the possible cells of the same mode or different mode or even different RANs to HO. Depending on the type of handover; intramode, intermode or intersystem, this information will be provided by BSs, ACS/SRRM or CoopRRM.
The inter-RAN and Intermode handover decision will be advised (or taken) by the following logical entities:

- HO between WINNER and legacy RAN, the decision will be advised by the CoopRRM
- HO between LA and (WA or MA) the BSwa or the BSma will be advised the decision (it is assume that the BSla will be controlled by the BSwa or the BSma)

- HO between WA and MA, the decision will advised by the ACS/SRRMW .

If we follow a self-organized and partially distributed approach the intramode handover decision could be taken by the BSs /UTs of the same mode, in a similar way to the current 802.11 standards, in which there is not a central entity over the BS as the current cellular systems. The BSs, of the same mode in the same deployment zone, could use a protocol to exchange control messages between them, in a similar way that the 802.11 APs use the IAPP protocol, to give a continuous coverage to support terminal mobility.

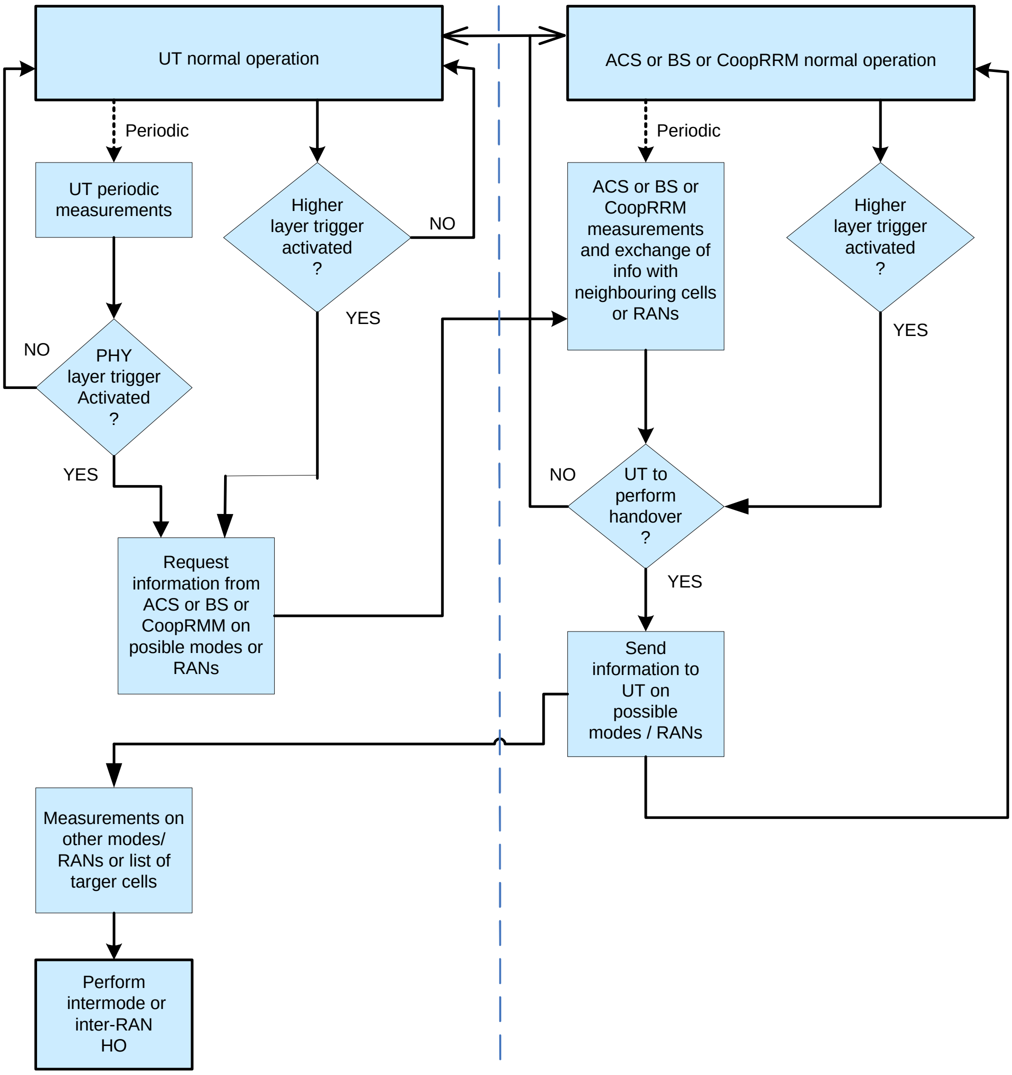

Figure 2-2: WINNER integrate handover/triggers and measurement process (intra-mode, Intermode and inter-RAN)

2.2.1.1 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 active state, when a data flow is requested by the UT or the network, this data flow is mapped to one of the 18 service classes presented in section 2.2.3.2. These service classes could be served by 2 or 3 WINNER serving deployment scenarios (WA MA or LA). In the case of the UT would be served by an adequate deployment scenario, as default the UT will try handover to a neighbouring cell of the same mode when an intramode trigger is activated. Only in the case there were not cells of the same mode available, or when some specific intermode triggers are activated (see intermode handover section) the UT will handover to other mode or RAN (for example increase or decrease of UT velocity).

Triggers on current cell that initiate intramode HO:

- Signal strength
- Interference level
- BLER
- SINR (Signal -to-interference plus noise ratio)
- Cell congestion

All these triggers are based on information from the BS/RN and UT of WINNER system, without additional processing of other logical nodes, and therefore the HO decision could be very fast and efficient.
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

Hereafter, we propose first intramode algorithms, assuming that two different cells (Current and Target cells) can provide the QoS requirements related to the service requested by the user.
M is a metric for signal strength, the common pilots signal strength. Th refers to a given threshold.

Handover from Current cell to Target cell, based on coverage criteria (first step, for reference):

If $M_{Best,Current} < Th_{Current,1}$ Then measurements are triggered on target cells that are included in the neighbouring cell lists.
If $M_{Best,Current} < Th_{Current,2}$ and $M_{Best,Target} > Th_{Target}$ Then handover is performed to the best measured neighbouring cell.

Handover from Current cell to Target cell, based on load and coverage criteria:

If $M_{Best,Current} < Th_{Current,1}$ Then measurements are triggered on target cells that are included in the neighbouring
If $M_{Best,Current} < Th_{Current,2}$
    If $M_{celli,Target} > Th_{Target}$ and $Load_{Celli,\ Target} < Th_{load}$
    Then handover is performed to the best measured neighbouring cell.

It should be noted that the threshold for the load criteria can be absolute or relative, if it is possible to compare directly the load of two systems.

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.).

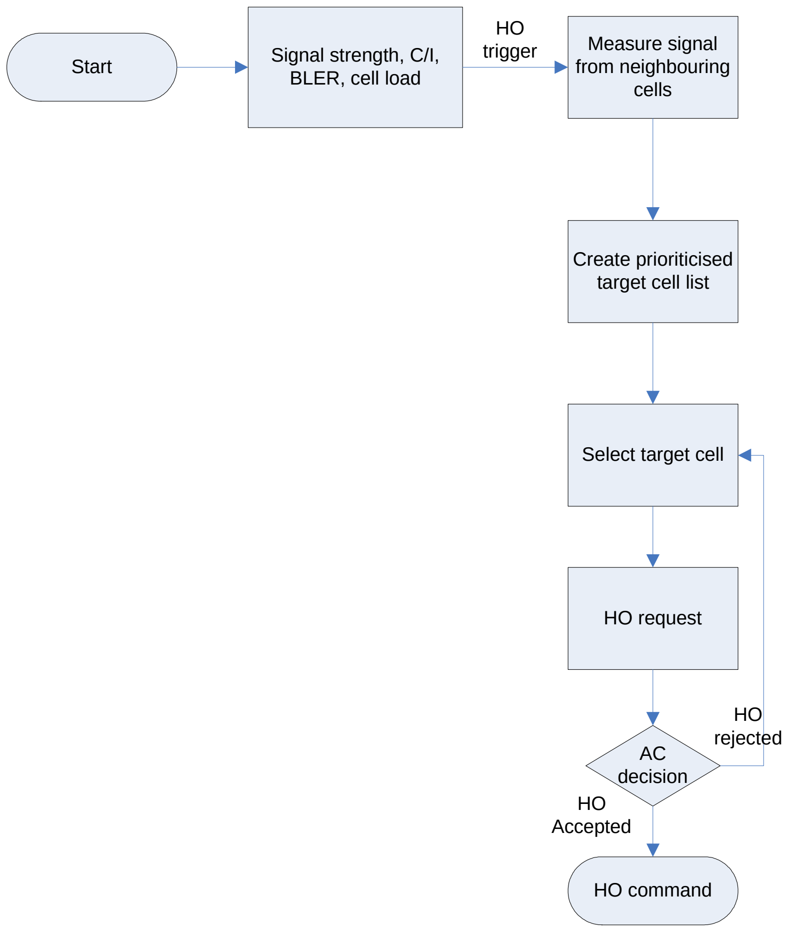

Figure 2-3: WINNER intramode handover flowchart.

In the previous figure, a general flowchart for the intramode handover algorithm is presented. The algorithm in the UT performs periodic measurements and, when a trigger for intramode handover is activated, then it is measured the signal included in the neighboring cells list and these are ranked as better candidates to handover. Then, it is selected as the first target cell, then the Admission Control decides to accept the user in the new cell or not and 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.

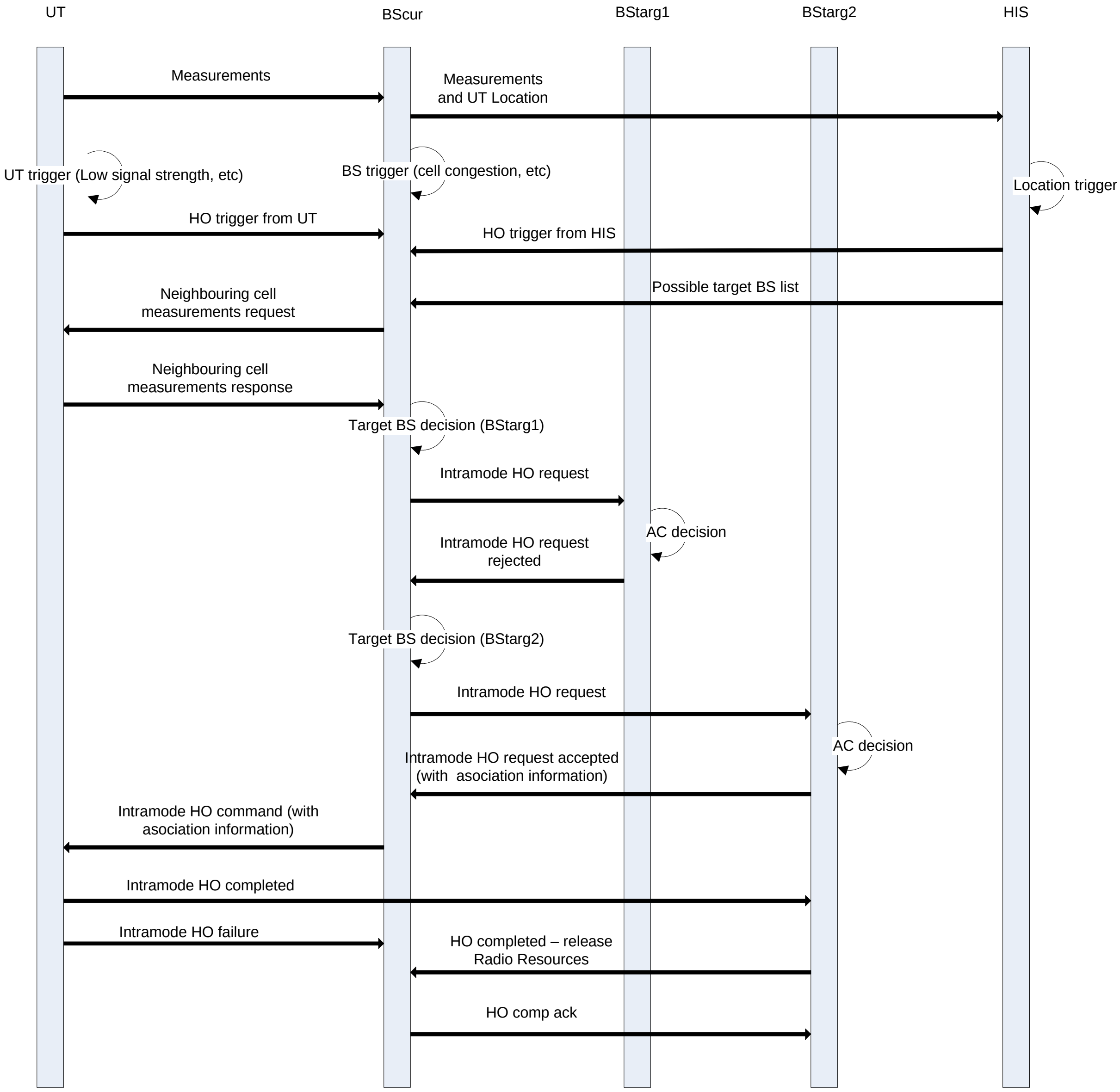

Figure 2-4: Example of WINNER intramode handover process

Figure 2-4 presents an example of an intramode handover process, presenting the signaling exchange between the different nodes. 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 it receives 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.

2.2.1.2 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.

In the case of handover from the wide area to the local area also there are specific triggers, e.g. the need for higher datarate services. In the case and in the case of handover from local area to wide area the most important trigger will be the UT velocity.
As it was explained in the intramode handover section, after the initial selection of the service class and the associated modes to an user data flow, the UT will be maintained in this mode, except in the case not available cell in this mode will be available or some changes would occur in the UT environment / or data flow would happen (the specific Intermode triggers). Figure 2-5 presents the Intermode flowchart following this principle. In the case of any WINNER cell will be available then the cell of other RANs would be checked, in this case the cell selecting process in the legacy RAN will be similar to intermode handover process.

Triggers for intermode handover (common with intramode handover):
- Signal strength (and not cell of the same mode available)
- Interference level (and not cell of the same mode available)
- BLER (and not cell of the same mode available)
- SINR (and not cell of the same mode available)
- Cell congestion (and not cell of the same mode available)

Specific Intermode triggers:
- Increase/decrease of UT velocity
- New service request/release
- Terminal location
- Current cell load is higher than target cell load
- User preferences (price, operator) can be fulfilled on target mode but not on current mode (or RAN)
- Higher datarate reachable on target mode than on current RAT (or RAN)
- Higher QoS reachable on target mode than on current mode (or RAN)
- User's class of service (bronze users on UMTS/ gold users on WINNER for instance)
- Operator's policy concerning service (voice on UMTS for instance) service availability
- QoS violation

In order to have efficient and fast handovers, we assume a decentralized and hierarchical mode control architecture, that is presented in chapter 5, in which the intermode handover decision is taken by the Bswa/ma (which is the base station of the wide area mode) that controls several BSla ((which is the base station of the local area mode). In this approach the ACS role will be limited to coordinate a set of BSwa or Bsma.

Taking into account the BSma can control several RNs, it is foreseen that the BSma could control also several BSla like the Bswa. Therefore the BSma will have the same control functionality over the BSla than the Bswa.

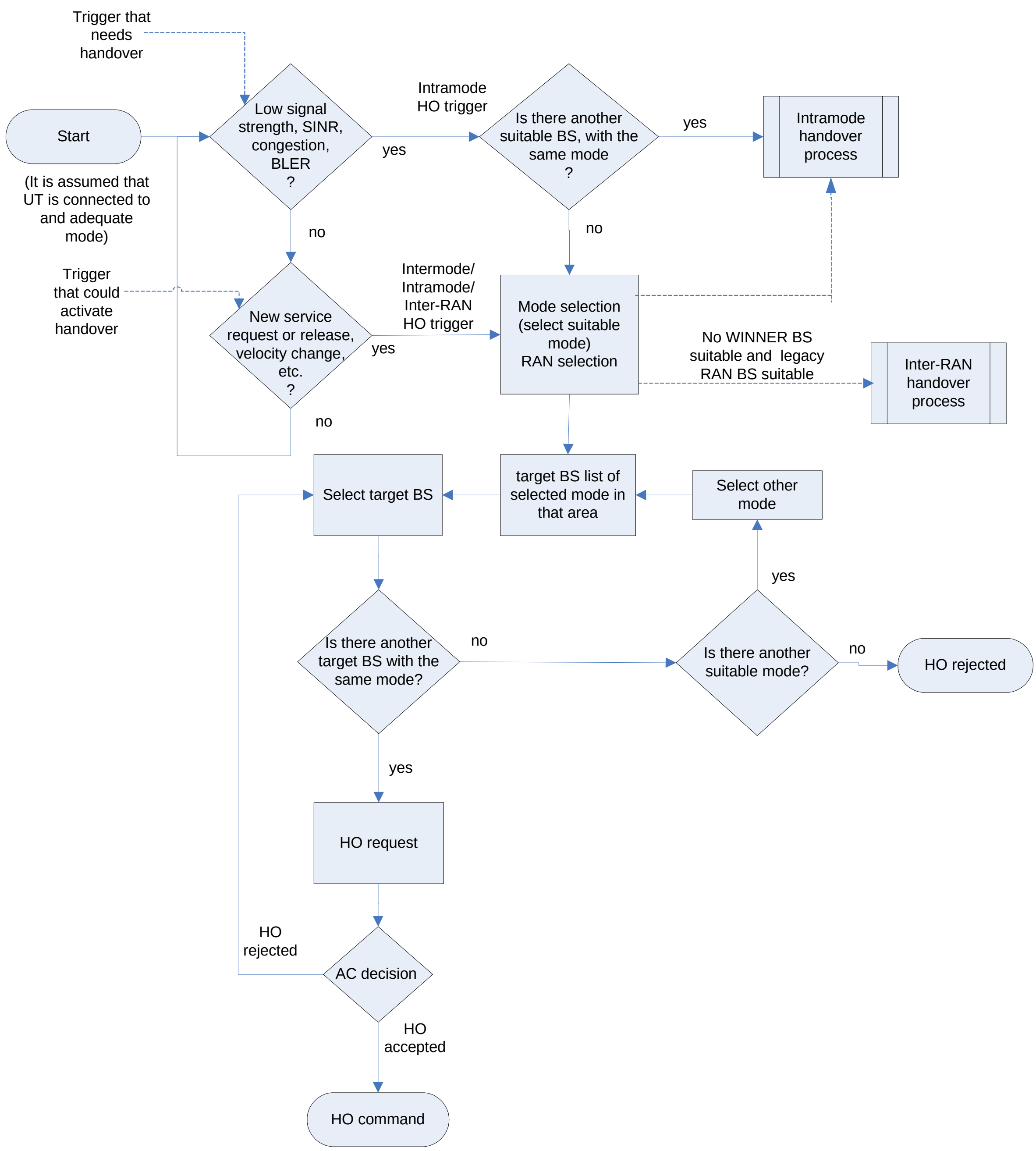

Figure 2-5: WINNER intermode handover flowchart (including its relationship with the intramode process)

In the previous figure the basic flowchart for an intermode handover is presented. As we can see the bad signal strength and quality (BER) and the cell congestion triggers the intramode handover process, and then it is search a BS in the same mode to handover to, if not BS in the same mode is found, then a Intermode handover is triggered. There are specific triggers that directly activate Intermode handover as new services request/release and velocity changes.

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 it is created a list of target cells and it is checked with the admission control to which cell the user can be admitted.

An example of the signaling process for an intermode handover from local area to metropolitan area (coordinated by a BS of the wide area) is presented in the following figure.

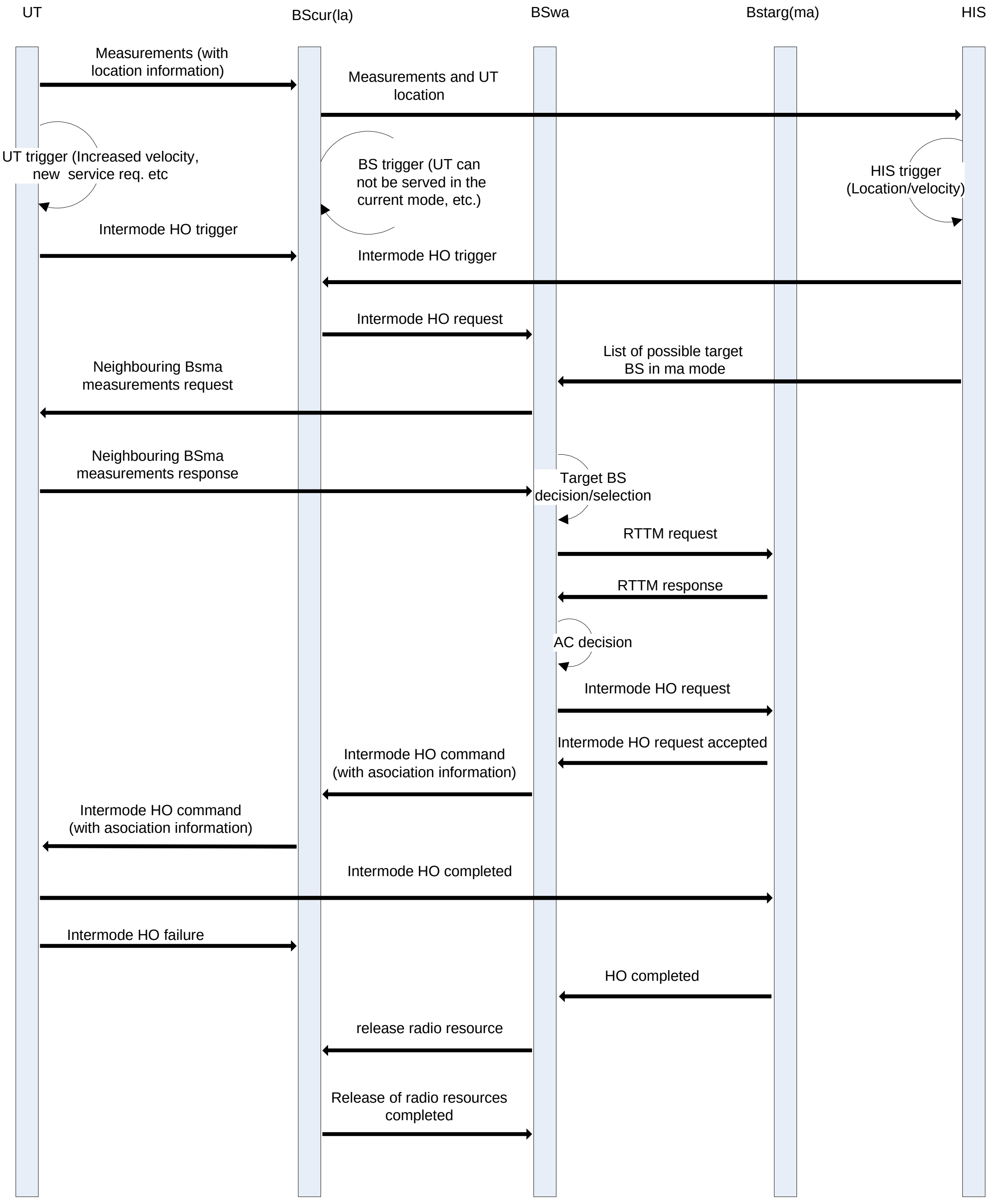

Figure 2-6: Example of WINNER intermode handover process from LA to MA coordinated by BSwa

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.

If this intermode handover procedure is being handled by the BSma, then the procedure should be like the one in figure 2-5.

2.2.1.3 Handover from wide area cell to local area cell

This handover could be initiated either by the UT or the BSla. However, the decision for the handover acceptance will be made by the Bswa/ma. The 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 PH layer mode) are the following:

- Need for higher datarate provided by the local area cell (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 conclude 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.

2.2.1.4 Handover algorithm from local area cell to wide area cell

This handover could be initiated either by the UT or the BSla. However, the decision for the handover acceptance will be made by the BSwa. Example 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, before the trigger of the handover, there is 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 a pre-trigger is activated 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.

2.2.1.5 Improvement of Intermode handover using neighbouring cell list

As described in the section above, the handover process could be improved with the use of neighbouring cell lists information. In the case of intermode handover we deal with cells of different size, the size of the cells of the modes is not the same, i.e. the wide area mode has very big cells though the local area mode has very small cells. Since there are limitations to the terminal and the signalling of the network, the neighbouring cell lists should be kept as low as possible.

Useful neighbouring cell lists can be defined as follows: for a given User Terminal in a given cell, that wants to perform inter-mode handover, a neighbouring cell is useful if its coverage area is such that the User Terminal could be served by this cell during a minimum amount of time. This definition enables to avoid resurgences. It is only based on coverage criteria, in order to be independent of the user's service or QoS requirements.

3 types of neighbouring cell list sizes can be distinguished:

- Low size: if the primary mode has cells with far lower coverage than the cells of the target mode.
- Medium size: if the primary mode has cells of the same coverage order as the cells of the target mode.
- High size: if the primary mode has cells with far higher coverage than the cells of the target mode.

The main issue is on the method used to inform the User Terminal of the identity of the neighbouring cells that it shall measure for handovers. Neighbouring cell lists can either be broadcasted, at cell level, or sent directly to the User Terminal via dedicated messages. If they are broadcasted, then all the User Terminals served by a cell will measure the same neighbouring cells. This can be used if neighbouring cell lists are of low size, e.g. if the target cell(s) cover the whole primary cell's area. However, if the primary cell covers too many candidate target cells, then dedicated messages are needed, for each User Terminal, to indicate specific neighbouring cell lists corresponding to that User Terminal's position.

SRRM could be the responsible entity for computing neighbouring cell lists for WINNER inter-mode handover. Location information could also be an input to build neighbouring cell lists at mobile terminal level, when dedicated neighbouring cell lists are required. Besides, neighbouring cell lists' size could be adapted depending on the current situation: they could be reduced if the User Terminal is in an emergency situation and needs to perform inter-mode handover as soon as possible, or they could be increased (or kept at broadcast level) if the User Terminal has the capability to perform inter-mode measurements without degradation.

2.2.1.6 Information exchange requirements for WINNER intermode handover

To enable smooth intermode handover some information is required to be exchanged by the BSs of the WINNER modes. An approach could be that every Bswa/ma should send all its cell information and updates at the WINNER ACS/SRMM and the BSla send its cell information to the associated BSwa/BSma. Therefore all decisions for intermode HO will be made by the ACS or the BSwa/BSma,. Furthermore, some of the information might be also needed to be broadcasted.

Required Information from the local area BS (BSla):

- Number of RNs attached to BSla (in case of mobile RNs the number of attached RNs is expected to change)
- Identification of RNs by a unique number, functionality: fixed/mobile, L1/L2/L3, conventional/cooperative
- Max number of hops
- Location within the wide area cell (CELLwa, if available)
- Range of the short range cell (CELLla) (this can be change in case of mobile relays leaving/joining)
- CELLla information: cell id, current load, max load, power information
- Handover statistics per RN and BSla (successful, drop rate etc)

Information from the wide area BS (BSwa):

- CELLwa information: cell id, current load, max load, power information
- Handover statistics

Information from MT

- Measurements: signal strength, interference, BER/PER, velocity etc
- User preferences: QoS, operator, cost etc.

2.2.1.7 Intermode handover employing HIS

In this chapter we investigate WINNER intermode Handover by employing the HIS system. To further increase overall system performance resources should be disburdened as much as possible. One way to achieve this is by increasing the cell size. The HIS (reference) is able to identify cell borders in arbitrary environments. Mobile terminals permanently report their current position and link state information to the attached databases. This combined information can be employed to enable mobile terminals currently outside the cell to communicate with the other points of attachment (e.g. BSs, relay nodes) at the cell border.

This approach is illustrated in Figure 2-7. Monitoring positions of all terminals within connected cells and at different WINNER modes (WMi), the HIS is able to identify different attachments points at the cell borders, which are capable to support different communication. This would increase the coverage area significantly and improve the handover procedure by reducing the latency.

2.2.1.7.1 Identification of point of attachment position

It is possible to identify cell borders by evaluation of location related measurement reports. Mobile terminals that are located near cell borders and that do not move too fast can thus be identified easily by the Hybrid Information System. These mobile terminals may in principle be used by other mobile terminals currently outside the cell coverage to enable communication to the other WINNER modes.

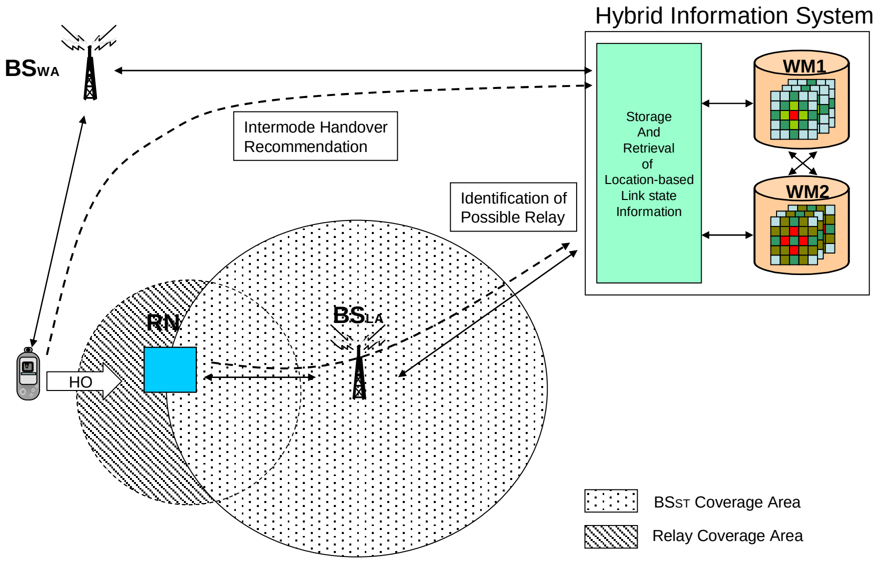

Figure 2-7: WINNER integrate handover/triggers and measurement process (intra-mode, Intermode and inter-RAN)

The Hybrid Information System could inform mobile terminals that are leaving the cell about a possible topology within their vicinity to establish communication with that attachment point. In that way measurement reports for areas outside the cell coverage may be gathered. These reports may then be used to determine the link quality outside the cell coverage to recommend possibly modes and hence possible attachment points to arriving mobile terminals. It should be noted that link quality measurements lose much of their relevance with varying position information. To solve this problem in general, each measurement needs to be associated not only by the mobile terminal’s position but also with the current position of different attachment points including fixed and mobile relays. To overcome increase in data complexity, probabilistic statements based on measurements and basic assumptions for the signal range of currently available points can be considered.

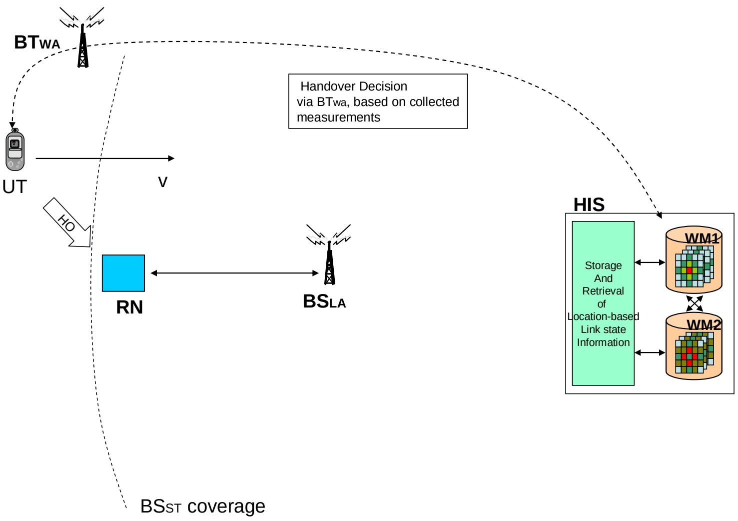

Figure 2-8: HO Trigger in different deployment scenarios

2.2.1.7.2 HIS in a multi-hop scenario

To improve the system performance, the coverage detection procedure described above should take into consideration the switching points between different types of attachments points. For example to switch between the relay nodes (fixed or mobile) and the access point. Here the switching point take into consideration not only the distance from relay to the base station, but also several other inputs. For instance the relay-access point link will benefit from directional antennas employed at both ends, so the antenna gain is an important factor for the link capacity. Different criteria for relay – access point handoff might be considered that optimize end-to-end throughput or delay. Furthermore the number of mobile terminals connected to the relay – which primarily determine the needed bandwidth – is crucial because the relay - access point link has limited capacity.

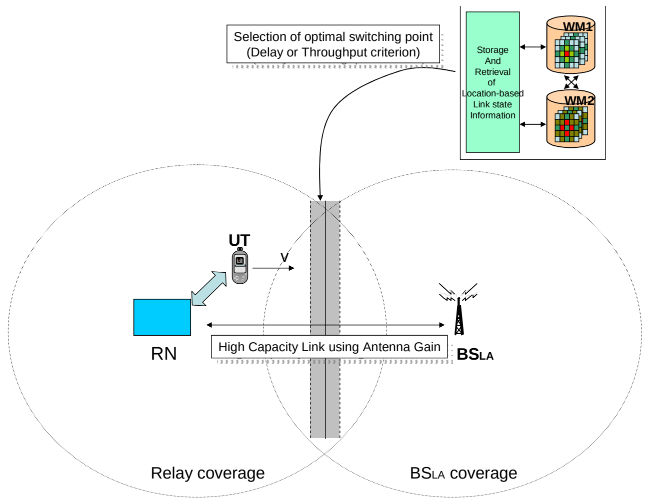

Figure 2-9: Optimization of Relay - Access Point Handoff Point

Figure 2-9 illustrates the handoff optimization problem. A mobile terminal that maintains a communication link with the relay moves towards the base station area of coverage. The Hybrid Information System needs to rely on measurements formerly reported by other terminals to find an optimal handoff point according to some criterion. This decision will not only rely on coverage information derived from received signal power measurements but will rely on several inputs such as the user's QoS requirements (to determine delay or throughput optimization), current cell load (both for relay and base station) and mobile terminal velocity and movement direction.

Combination of several input parameters for trigger generation is need. The following are examples for how to find the optimal handover points between relay and base stations:

- **Frame Measurement Report**

  Frame measurements can be used to easily gather Received Power Histograms (RPI) for different source stations. Thereby measurements are only made during frame transmissions and are each associated with the MAC address of the frame source. This is especially useful to perform measurements for a multiple stations simultaneously. If relay and access point operate on the same channel both signal strengths can be measured simultaneously.

- **Channel Load Report**
  This measurement both takes into account the physical carrier sense mechanisms (Clear Channel Assessment - CCA) as well as the virtual carrier-sense mechanisms (Network Allocation Vector - NAV), to determine the current channel utilization. This is especially useful to estimate the available system capacity.

- **Medium Sensing Time Report**
  The Channel Load Report gives information on the current channel utilization. The Medium Sensing Time Report aims to give more detailed information by not only indicating a percentage of the utilized channel but reporting a histogram of sensing times, which allows for a more sophisticated view on the channel status and can give hints to estimate the packet delay to expect.

- **STA Statistics Report**
  This measurement can be used to query different counters within the mobile terminals, such as retry, multiple retry and failed counters. This allows for information gathering that does not only account for the physical layer but incorporates the link layer.

2.2.1.8 Fuzzy logic based mobility management

The big number of parameters to be taken into account in the sophisticated mobility management algorithms, for system B3G systems, it is recommended to use advanced mathematical methods to jointly compute these variables, one of the most promising proposals is the use of fuzzy logic.
In this section, the focus is on the mobility management between WINNER modes. The handover scheme using fuzzy logic studied further in section 3.3.1 is applied. The aforementioned handover model is generic and suitable to WINNER inter mode context as well. Mobility management using fuzzy logic is governed by a set of fuzzy rules. We only need to adapt rules to WINNER modes characteristics.
Hereafter, an example of fuzzy rule adapted to handover between WINNER modes.

IF *(Current system = WINNER_FDD)* ***AND*** *(Mobile terminal velocity = LOW)* ***AND*** *(WINNER_TDD coverage = MEDIUM* ***OR*** *HIGH) AND (WINNER_TDD load = LOW* ***OR*** *MEDIUM)* ***THEN*** *(handover to WINNER_TDD)*

This rule deals with situations where the UT is attached to WINNER_FDD and it is moving with a low velocity. If WINNER_TDD offers a satisfactory coverage and load levels, then a handover to WINNER_TDD is decided.

2.2.2 Congestion Avoidance Control

Wireless networks have limited radio resources and should be managed very carefully so that they operate in normal situations, thus assuring that the users will receive the requested Quality of Service for their requested applications/services. Because of the limited radio resources, there have been defined Radio Resource strategies in order not only to assure the QoS guarantees to the users, but also to assure that the user don’t violate these agreements from their side. When, for example, there are too many users admitted and the users don’t receive the agreed QoS, then the network is overloaded and this state of the network is defined as congestion situation. Congestion situations are very dangerous for the network because they cause many problems, such as increased interference, loss of packets, low bandwidth availability and from the user’s point of view it causes decreased quality in the service reception which leads to the user’s disappointment.

The congestion situations can be avoided applying a set of techniques which work on this direction. These techniques set a mechanism defined as “Congestion Avoidance Control” mechanism, which has the important task of controlling the load of the network by restricting the admission of new user sessions and resolving unwanted overload situations. The Congestion Avoidance Control mechanism consists of two mechanisms:

- Admission control, which is the mechanism that receives the requests for new sessions (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 Control, which is the mechanism that controls the load of the network by continuously monitoring the load and acting if the load exceeds some pre-defined thresholds.

2.2.2.1 Admission Control

Admission Control is one of the key RRM mechanisms that ensure the good operation of a network, by admitting or rejecting new user requests based on criteria such as the load of the network. In general, the Admission Control mechanism ensures that the admittance of a new flow into a resource constrained network does not violate the service commitments made by the network to already admitted flows. The admission control schemes are the decision making part of networks with the objective of providing to users services with guaranteed quality in order to reduce the network congestion and call dropping probabilities and achieve as much as possible resource utilization. In current legacy systems the admission control only examines the requests for entering the network in a specific cell. In the WINNER system, which will have different operational modes and will cooperate with legacy networks the admission control is a more complicated mechanism. The algorithm will not only examine admitting the user into a specific cell, but it should select the most suitable WINNER mode for the service he requests and if the WINNER system is overloaded, it should find the most suitable legacy system to serve the user.

For the cooperation amongst WINNER modes the admission control has the task to handle the new user requests that can be handled inside the WINNER system and the intramode and intermode handover requests. The decision about admitting or rejecting a new user request is based on certain criteria. These criteria were defined in WINNER I for the cooperation between WINNER and legacy systems and can be more or less applied also for the cooperation amongst the WINNER modes. These criteria are based on information such as the load of the target cell/mode after the admission of the new request, statistics of current flows, service requested from the user, priority of the flow (based on QoS criteria), resource restrictions and physical layer measurements. It is assumed that flow statistics will include:

- throughput per flow, in previous super-frame
- time-frequency-spatial resource use per flow
- spare capacity within cell in previous super-frame
- path loss estimate, and
- interference estimates

For the admission control inside the WINNER system, very important criteria for selecting the target mode for the user to be accepted will be the mobility and the throughput requirements of the requested service. For admitting a user in a short area cell, the user should have very low velocity and/or request very high datarate. For admitting a user in a wide area cell the user should have very high velocity, but also low datarate requirements. In the case when a user has high mobility and high datarate requirements then the decision will be very critical, since admitting the user in a short area cell will result into many inter-cell handovers creating increased load in many short area cells and also a lot of signalling load used for the inter-cell handovers. On the other hand, admitting the user it a wide area cell will cause problems either to the load of the wide area cell (it will increase very much since the user will have high load due to the high requested datarate) or to the user (he will received the requested service with decreased quality due to the insufficient datarate). Thus finding the suitable mode for the user is a very critical decision. This is ongoing activity and results will be given in the next deliverables.

If the resources in the target cell/mode are sufficient, then the user request is accepted. On the other hand, if the resources in the candidate cell/mode are limited, a number of different actions (not by this order) may be taken taking into account the characteristics of the user’s request (handover / new request, priority etc.):

- reduce requests for connection/flow in question and/or for lower priority flows
- resource re-partitioning/renegotiation
- lower load in (interference from) neighbouring cells
- handover flows (cell/mode/RAN)
- drop flows.

In WINNER I it was defined and assessed an algorithm to support the cooperation between WINNER and legacy systems. This algorithm can be adapted for the cooperation amongst WINNER modes. Here we will describe in general the concept and the steps of the algorithm that will be described in more details in the next deliverables.

The algorithm takes into account all the information available for the user, such as mobility, location, and the characteristics of his request such as service characteristics and the user's priority. The general steps that the algorithm will follow are the following:
The algorithm makes a list of the candidate serving modes and candidate cells for each mode based on location information and service requirements such as throughput. The lists contain the candidate modes and cells capable of providing the requested flow service and they are ordered in such a way that better fulfils the service/user requirements.

- After selecting the target mode, the algorithm should also find the target cell based on location information and coverage capabilities.
- Then the algorithm checks the priority of the flow (if it is new or from handover) and the number of flows in the queue (if the flow is new the flows in the queue should be served first otherwise the priority plays the important role).
- The algorithm checks if there are sufficient resources in the target cell/mode and if so the flow is accepted. If not then there are two options; check if another cell/mode can serve the user and has sufficient resources or perform resource re-partitioning/renegotiation in the target cell/mode, if the priority of the user is relatively high.
- If none of these actions can gain the needed resources for admitting the flow and the flow has relatively high priority then it is enters the admission queue in a position in front of lower priority flows.
- The flow will remain in the queue until:
    - The needed resources become available (i.e. by completion of other flows)
    - The flow leaves the cell, i.e. the user moves to another cell, or the flow is completed
    - The flow is terminated due to timeout.

The admission control algorithm for the cooperation amongst WINNER modes aims to maximize the number of admitted or in-flow traffic sources supported over the WINNER system, while guaranteeing their QoS requirements and ensuring that the new connection does not affect the QoS of the ongoing connections.

2.2.2.2 Load Control

Load control is another part of the congestion avoidance mechanism. The Load Control is one of the key RRM mechanisms that ensures the good operation of a wireless network, keeping the load of the network in normal boundaries, performing traffic balancing between modes (in the WINNER system) or cells of the same mode preventing congestion situations. Reactive load control is used to face a situation where the networks are in an overload situation and the users' QoS is at risk due to increase of the interference, mobility aspects, low bandwidth availability etc. When already admitted users cannot satisfy their guaranteed QoS to their services, for a specific percentage of time, then the network is considered to be in an overload/congestion situation. Moreover, in the unwanted case when the congestion situation occurs, the load control performs several actions to decrease the amount of traffic in the congested cell/mode. The load control also monitors continuously the system to prevent these unwanted congestion situations.

The load control algorithm is triggered in order to ensure that the system is in stable state. It calculates the predicted load per cell/mode based on information such as load prediction, statistics of current flows and resource restrictions in the target cell/mode. In particular:

- The preventive load control takes care so that the network does not get overloaded and remains stable and attempts to improve the system performance by distributing users/sessions/resources among modes/cells/sectors.
- The reactive load control attempts to bring the load back to stable condition as fast as possible
- The load sharing aims to distribute the load (the offered traffic) and/or resources between "resource owners" (BSs and RNs) such that the resources are efficiently utilized

As it is also mentioned above, despite the fact that the load control tries to prevent congestion situations, the network could be overloaded for some reasons (i.e. users violate service agreements etc.). At this situation several actions should be taken in order to decongest the cells/modes. Potential actions for a "congested cell/mode" are:

- interaction with the *service level controller* and restriction of incoming traffic
- interaction with the admission control and denial of new flow requests
- change of TDD asymmetry factor
- reduction of requests (e.g. bit rates) for (usually low priority) flows (within limits as specified in the flow QoS specification)

- attainment of more resources by resource re-partitioning or lower load in (i.e. lower interference from) neighbouring cells
- handover of flows to another cell/mode
- dropping of low priority flow(s)

In WINNER I it was described in details a reactive load control algorithm for the cooperation between WINNER and legacy RANs. This algorithm although it was defined for the cooperation between WINNER and legacy RANs, it is generic enough so it can be adapted for the cooperation amongst WINNER modes. Here we will describe in general the concept and some generic steps of the algorithm that will be defined and described in more details in the next deliverables. The reactive load control algorithm is divided into three phases:

- *Detection phase*: The algorithm continuously monitors the network and periodically checks the load of the cells/modes in order to detect an overload situation in any of them. It is considered that a cell/mode is overloaded if the load factor is over a certain pre-defined threshold during a certain amount of time.
- *Resolution phase*: This is the phase that the algorithm is trying to resolve the problem that causes the overload situation. Some of the actions that the algorithm performs to resolve the congestions situation have been mentioned above.
- After resolving the congestion situation, the algorithm enters the *recovery phase*, where it tries to restore the QoS to the flows, which QoS was degraded in the resolution phase. A recovery algorithm is necessary, because the flows should restore their QoS to the agreed state in the QoS service agreements.

2.2.3 QoS based management

For the purpose of demonstrating the proposed approach to QoS management we assume that the different WINNER modes cover short range, metropolitan range, and wide range deployment areas. With other words, we assume that there are at least three physical modes of operation of the WINNER RAN associated with each deployment scenario. To these physical modes (i.e., also deployment scenarios), we associate service sets, which are a grouping of applications requiring different data rates, delays and ranges. The service sets are stored in a database located or associated with the ACS. In the following, we will refer to the term "applications" as service classes, which can be composed of several applications. Therefore, when we talk about application priority, it will in reality refer to the priority of the service class that the application belongs to. An exemplary classification considering the previous parameters, is shown in Table 2-2

Service class	**Required throughput**	**Tolerable delay**	**Mobility/Deployment Scenario**	**Default WINNER mode**
Real time (RT) collaboration and gaming	1–20 Mbps	<20 ms	Low/local area	#1
Geographic RT datacast	2–5 Mbps	<20 ms	Global/local area	#1
Simple interactive applications	64–512 Kbps	20–100 ms	Any	#3
Interactive high multimedia	2–5 Mbps	20–100 ms	Low/wide area	#2

Table 2-2: WINNER services and associated WINNER modes

For the QoS management we will need an extra functionality that would indicate the capabilities of the user's terminal and a default WINNER mode associated with the user's profile and typical services that can be requested and are associated with this profile. This information may vary from one connection to another, because each time the user will request an application, the default WINNER mode associated with this application (see application profile) will be stored. The preferred mode of the user will then be the mode the user tried to connect most often to. Information about the preferred mode must be stored in a small database located at the user terminal. This preferred mode is included in order to minimize the probability of IMHO when the user will requests his first application.

An additional type of information (dynamic information) will be required about the WINNER mode which the user belongs to or is connected to.

2.2.3.1 User profiles

Each connecting user will be associated with specific information composed of a personal user profile, stored in a database that is located at the UT. Among this information, some will be static and will remain constant during the whole connection, describing the general profile of the user; and some will be dynamic, describing the activity of the user on the network.

For a WINNER user the static information will include the following:

- User's ID: This is the identification of the user, among the different connected users.
- User's subscription profile: This includes information about the operator to which the user belongs and the type of contract the user has. This information will be used to know if the user can be granted access to the network and/or requested service. The user would be able to access a network if the network belongs either to his operator or to an operator that has agreement with his home operator. Information about, operators' agreement must be stored at the CoopRRM. Eventually, this information will be taken into account when defining the priority of the user.
- User's origin: This describes whether the user comes from a HO process initiated on another network or is a new user. A user coming from a HO session will have higher priority than a user requesting a new session.
- User's miscellaneous information: This was especially meant to describe an "emergency" user. An emergency user is someone that would be highly prioritized in any case thanks to professional matters (e.g. policemen, paramedics, firemen, ...)

Dynamic information for a WINNER user is:

- User's RAN. This is the RAN to which the user is connected.
- User's KPIs. These values are stored to keep trace of the QoS the user is provided with.
- User's ongoing applications. This sums up the resources consumption of the user. If he runs no application, his consumption is zero.
- User's request for application. In case the user is requesting a new application, this is to know what his new resources needs are.
- User's location. This information is used to keep trace of the user's position. By comparing changes in positions, it can also determine the user's mobility level.
- User's basic priority: This will represent the priority level of the user during the time he will be connected to the network, taking into account all the previously mentioned priority related information. The user priority levels are then chosen based on the described reasoning.

Table 2-3 gives the chosen user priority levels. We must note that this is a simplified way to prioritize the users that was assumed for simplifying the simulation work. However, in reality, the users profiles and the priority levels will depend on a larger number of parameters (mobility patterns, details of the subscription profile, terminal capabilities, etc).

Contract type	User's origin	User's Basic Priority level
Emergency	Any	1
Type 1	HO	2
	New	3
Type 2	HO	4
	New	5
Type 3	HO	6
	New	7

Table 2-3: User profiles and chosen priority levels

2.2.3.2 Service sets

The service sets are a grouping of applications requiring different data rates, delays and ranges [Ref D1.4 of Winner I]]. The service sets information can be stored in a database located at the ACS. In the following, we will refer to the term "applications" as service classes, which can be composed of several applications.

Therefore, when we talk about application priority, it will in reality refer to the priority of the service class the application belongs to. The classifications considering the previous parameters is shown in Table 2-4

Service Class	Required Rate	Tolerable Delay	Required Mobility/Range	Default WINNER Mode
1. Real Time Collaboration and gaming	1-20 Mbps	<20 ms	LM/SR	#1
2. Geographic real time datacast	2-5 Mbps	<20 ms	GM/SR	#1
3. Short Control messages and signalling	8-64 kbps	20-100 ms	Any	#3
4. Simple interactive applications	64-512 kbps	20-100 ms	Any	#3
5. Interactive high multimedia	2-5 Mbps	20-100 ms	LM/GR	#2
6. Geographic interactive multimedia broadcast	2-5 Mbps	20-100 ms	GM/SR	#2
7. Interactive ultra high multimedia	1-50 Mbps	20-100 ms	LM/SR	#1
8. Simple telephony and messaging	8-64 kbps	100-200 ms	Any	#3
9. Data and media telephony	64-512 kbps	100-200 ms	Any	#3
10. Geographic datacast	64-512 kbps	100-200 ms	GM/SR	#2
11. Rich data and media telephony	2-5 Mbps	100-200 ms	Any	#2
12. LAN access and file services	0,5-50 Mbps	100-200 ms	LM/SR	#1
13. Multimedia messaging	8-64 kbps	>200 ms	Any	#3
14. Lightweight browsing	64-512 kbps	>200 ms	Any	#3
15. File exchange	Up to 5 Mbps	>200 ms	LM/GR	#2
16. Video streaming	5 Mbps	>200 ms	Any	#2
17. High quality video streaming	2-30 Mbps	>200 ms	LM/GR	#1
18. Large files exchange	1- 50 Mbps	>200 ms	LM/GR	#1

Table 2-4: Service classes requirements and associated WINNER mode

We assign the terms WINNER mode #1 corresponds to short range deployment scenario, #2 to metropolitan range scenario and #3 to wide range scenario, and the classification was performed taking into account data rates and range/mobility.
It is as if we had three sub networks for WINNER, and depending on the application chosen by the user, we will place him in one of these sub networks. We define a Default WINNER mode corresponding to the mode fitting the best with a given application. However, a same application can be provided by several RANs as shown in Table 2-5 which sorts out the service classes depending on the RAN(s) they are compatible with. The red colour highlights the default WINNER mode defined in Table 2-4.

Service Class	GPRS	UMTS	WLAN	WINNER #1	WINNER #2	WINNER #3
1. Real Time Collaboration and gaming				x		
2. Geographic real time datacast				x		
3. Short Control messages and signalling		x		x	x	x
4. Simple interactive applications		x		x	x	x
5. Interactive high multimedia				x	x	
6. Geographic interactive multimedia broadcast		x		x	x	
7. Interactive ultra high multimedia			x	x		
8. Simple telephony and messaging	x	x		x	x	x
9. Data and media telephony		x		x	x	x
10. Geographic datacast		x		x	x	
11. Rich data and media telephony				x	x	
12. LAN access and file services			x	x		
13. Multimedia messaging	x	x		x	x	x
14. Lightweight browsing		x		x	x	x
15. File exchange				x	x	
16. Video streaming				x	x	
17. High quality video streaming				x		
18. Large files exchange				x		

Table 2-5: Service classes and RAN compatibility

2.2.3.3 Application profiles

Each application requested by a user will be associated with specific information forming a unique application profile. This information will include the following entries:

- Application ID. This is the identification of the application. It will be created by concatenating the user ID and the application rank of request. For example, if the user U1 requests a fifth application, this application's ID will be U1_5. This way, for a user connection, each application will have a unique ID.
- Application service class. By knowing to which service class the application belongs, it is possible to determine its requirements in terms of rate, delay, mobility and range, and then to identify the resources needed for its operation. The service class of the application b is also used to identify the WINNER default mode of operation, the compatible WINNER modes and the compatible legacy RANs that can host the application .

- Application's priority. This will represent the priority level of the application, depending on its level of interactivity and rate requirement. The priorities are shown in Table 2-6

<table>
<tr><th>Service Class</th><th>Interactivity</th><th>Rate category</th><th>Priority</th></tr>
<tr><td>1. Real Time Collaboration and gaming</td><td rowspan="2">Highly interactive</td><td rowspan="2">1-20 Mbps</td><td rowspan="2">1</td></tr>
<tr><td>2. Geographic real time datacast</td></tr>
<tr><td>3. Short Control messages and signalling</td><td rowspan="2">Interactive</td><td rowspan="2">8-512 kbps</td><td rowspan="2">3</td></tr>
<tr><td>4. Simple interactive applications</td></tr>
<tr><td>5. Interactive high multimedia</td><td rowspan="3">Interactive</td><td rowspan="3">1-50 Mbps</td><td rowspan="3">2</td></tr>
<tr><td>6. Geographic interactive multimedia broadcast</td></tr>
<tr><td>7. Interactive ultra high multimedia</td></tr>
<tr><td>8. Simple telephony and messaging</td><td rowspan="3">Conversational</td><td rowspan="3">8-512 kbps</td><td rowspan="3">5</td></tr>
<tr><td>9. Data and media telephony</td></tr>
<tr><td>10. Geographic datacast</td></tr>
<tr><td>11. Rich data and media telephony</td><td rowspan="2">Conversational</td><td rowspan="2">1-50 Mbps</td><td rowspan="2">4</td></tr>
<tr><td>12. LAN access and file services</td></tr>
<tr><td>13. Multimedia messaging</td><td rowspan="6">Few seconds delay tolerant</td><td rowspan="6">8 kbps - 50 Mbps</td><td rowspan="6">6</td></tr>
<tr><td>14. Lightweight browsing</td></tr>
<tr><td>15. File exchange</td></tr>
<tr><td>16. Video streaming</td></tr>
<tr><td>17. High quality video streaming</td></tr>
<tr><td>18. Large files exchange</td></tr>
</table>

Table2-6: Classification of application priorities

The applications priorities are determined taking into account the needs in terms of interactivity, that is to say the maximum acceptable delay, and the resources in terms of throughput. Indeed, users running high datarates applications use more resources. In case of congestion, as it will be, the congestion resolution process will perform at first handover for high priority users. For simplicity reasons we assume here that users requesting high datarates applications are high priority users. When these users leave, the load decreases more quickly than if low datarates users were redirected. The application profiles will be stored at the ACS.
The dynamic information is the expected download time which is duration for which we expect the user to use an application, depending on the type of application and of the size of the file to be transferred.

2.2.3.4 Prioritization

This section gives an exemplary prioritization that was necessary for the purpose of investigating the performance of the QoS-management algorithm. In the real case, prioritization will be a more complex process that will include a variety of characteristics related to a user or application.
For the purpose of designing a simplified scenario for the performance investigation we adopt a two-level prioritisation in the QoS algorithms, taking into account both previously introduced user and application priority levels. The prioritization process will consider that the user priority is more important than the application priority. To describe mathematically this relationship, each user level is associated with a ten value, and each application level is associated with a unit value. By adding the two values, we get a total value that describes the global prioritisation level of a known application used by a known user. In case two global levels are equal, the first arrived is first served.

Global priority levels must be stored in a small database located at the ACS. They are sorted in an ordered table from high to low priority that is to say from the lowest global value to the highest. This priority table is updated each time a user requests a new application, during the admission control scheduling process. Examples of such application priority values are shown in Table 2-7 and Table 2-8.

Contract type	User's origin	User's Basic Priority level	User's Basic Priority value
Emergency	Any	1	10
Type 1	HO	2	20
	New	3	30
Type 2	HO	4	40
	New	5	50
Type 3	HO	6	60
	New	7	70

High

Low

Table 2-7: Priority values of the users

Service Class	Priority level	Priority value
1. Real Time Collaboration and gaming	1	1
2. Geographic real time datacast		
3. Short Control messages and signalling	3	3
4. Simple interactive applications		
5. Interactive high multimedia	2	2
6. Geographic interactive multimedia broadcast		
7. Interactive ultra high multimedia		
8. Simple telephony and messaging	5	5
9. Data and media telephony		
10. Geographic datacast		
11. Rich data and media telephony	4	4
12. LAN access and file services		
13. Multimedia messaging	6	6
14. Lightweight browsing		
15. File exchange		
16. Video streaming		
17. High quality video streaming		
18. Large files exchange		

High

Low

Table 2-8: Priority values of the applications

For example, a user having a type 2 contract is coming from a handover requests an application from the Service Class 11. His basic priority level is 4, so his basic priority value is 40; the application priority value is 4. His global priority value is hence 44.

In the following, we distinguish between user-oriented and network-oriented algorithms. User oriented algorithms specifically guarantee the user's QoS, whereas network-oriented algorithms are used to avoid or solve congestion situations, and to improve on the network's performance. From the user's point of view, each session is initiated with the connection process and terminated by the disconnection process, in case connection was successful. Between connection and disconnection processes, the user can request as many applications as the user terminal allows it; each application request is handled independently by the admission control process. In all these algorithms, a scheduling process was introduced to avoid problems resulting from massive users' arrival, and to prioritise users. Independently from these user-oriented processes, the network-oriented algorithms take place as background processes. Figure 2-10 presents the interactions between the different QoS algorithms. Let us consider the case of a generic user called User 1. User 1 gets connected to the WINNER RAN through the connection process; it is for example a user who turned his mobile phone on. After a certain time, he requests the access to an application, Application 1, which will trigger the admission control process, in order to determine whether and where the user can be accepted in the network. He requests another application called Application 2, goes through the admission control process again. Let us assume that Application 2 is a messaging type of application, and that Application 1 is a Video Streaming type of application. Application 2 ends hence before

Application 1. It can happen somehow that during the connection time of User 1, some congestion appears on the network he belongs to, which will activate the congestion control mechanism. For resources management purpose, User 1 will be redirected to another RAN thanks to the handover mechanism, where he will run his last application Application n. Then, he will leave the area and the disconnection process will terminate his session.

If the user has a WINNER access, his preferred mode, which is the default mode he connects to, is checked. If this mode cannot accept the user for network load reasons, the second preferred mode is checked, and so on and so forth until all modes have been checked. If none of the modes could accept the user, he is rejected. If he does not has a type of contract that gives him access to the B3G network, he is directed towards his preferred legacy RAN, which is the legacy RAN closest to the WINNER preferred mode. Similarly, all the preferred RANs are checked until the user is accepted. If none of them can accept the user, he is again rejected. When the user is accepted, he goes through the connection scheduling process, which organizes the users in a queue. The information concerning the preferred mode, the user's operator and WINNER belonging is contained in the user profile, located at the Mobile Terminal. The agreements between operators are checked at the CoopRRM. The SRRMs are responsible for the verification of the availability of the networks. This is shown in Figure 2-10.

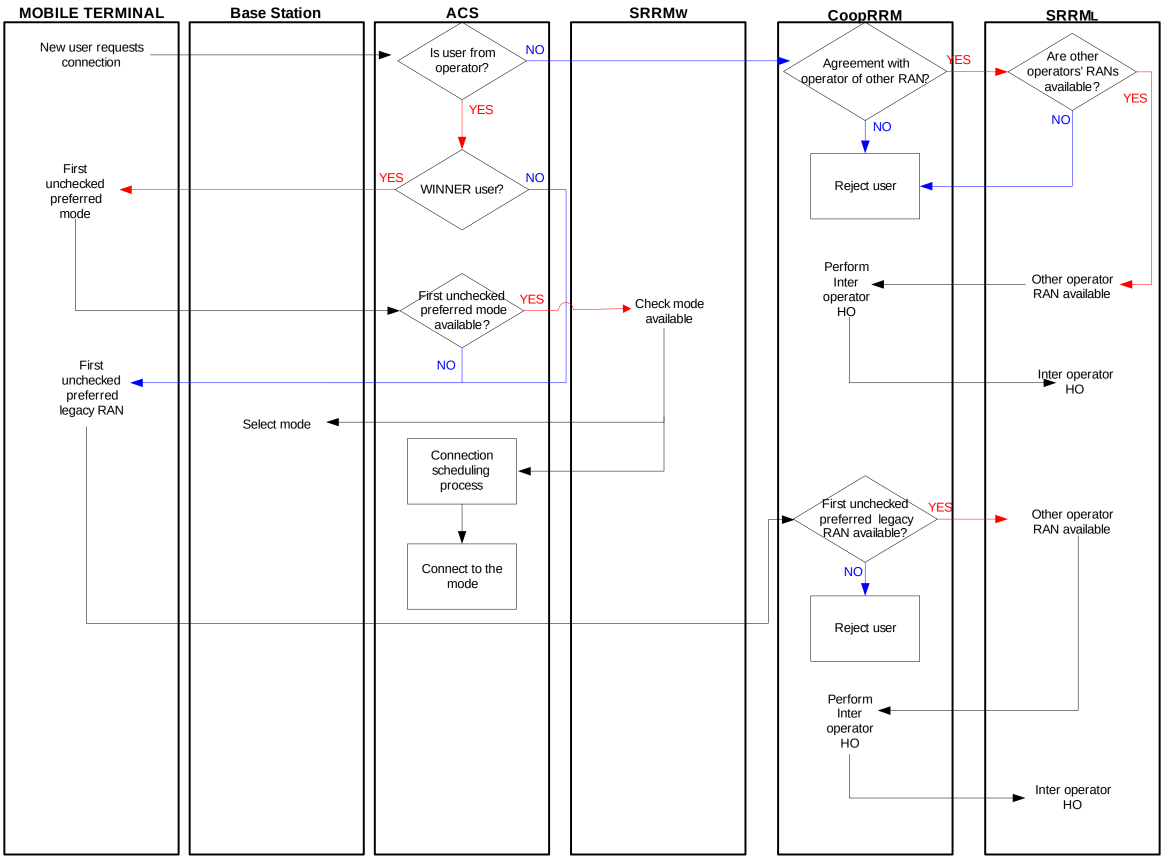

Figure 2-10: Message exchange during a connection process of a user

The SRRMW will make the decision of accepting or rejecting the user into the WINNER network, and the SRRML will follow the directives of the CoopRRM to perform inter-operator HO or ISHO. The CoopRRM will therefore eventually decide to accept the user, and then perform HO, or to reject him. Figure 2-11 describes the connection scheduling algorithm. This algorithm organizes and orders the incoming users to sort out the connections so that they do not all occur at the same time. When a new user is accepted, if the queuing buffer still has room, he will enter the queue and get effectively connected when all the users that arrived before him are connected, so that when he is in the "first position", and when the dead time is reached. This dead time corresponds to the duration after which a user gets connected to the network. Therefore, when the dead time has been reached, the first-positioned user connects to the network, and the time counter is re-initialized .Each time a user enters the network, an alarm is sent to the waiting users, which are those who could not get into the buffer because it was full. The alarm corresponds to the release of a place in the queue, which means that one waiting user can come in. This scheduling process does not take into account users' priorities, since we assume that the connection process does not necessitate many resources, so we apply a simple "First Come First Served" [WIND4.5] type of queuing.

The admission control is the second step in the QoS process. It is triggered when a user requests the use of an application. The QoS is ensured by two main elements: users are placed in the network that will best satisfy their requirements, and are prioritized taking into account both their basic user priority and the application priority as explained earlier. When a user requests an application, he has already gone through the connection process, so he was a priori already directed to a network he has access to. Since this algorithm is meant to work in the WINNER RAN, users that request an application are supposed to be already connected to a WINNER mode thanks to the connection algorithm.
Therefore, when the user has asked for a particular application, the first following step is to check what WINNER mode is the more suitable for this application. Two parameters are taken into account for this selection: first, the default WINNER mode, which is supposed to provide the best QoS in terms of rate and delay as described in the service sets; then, the mode that has the more chance of keeping the user until the end of the application, depending on the predicted download time versus the mode status. However, we have not implemented practically this alternative.

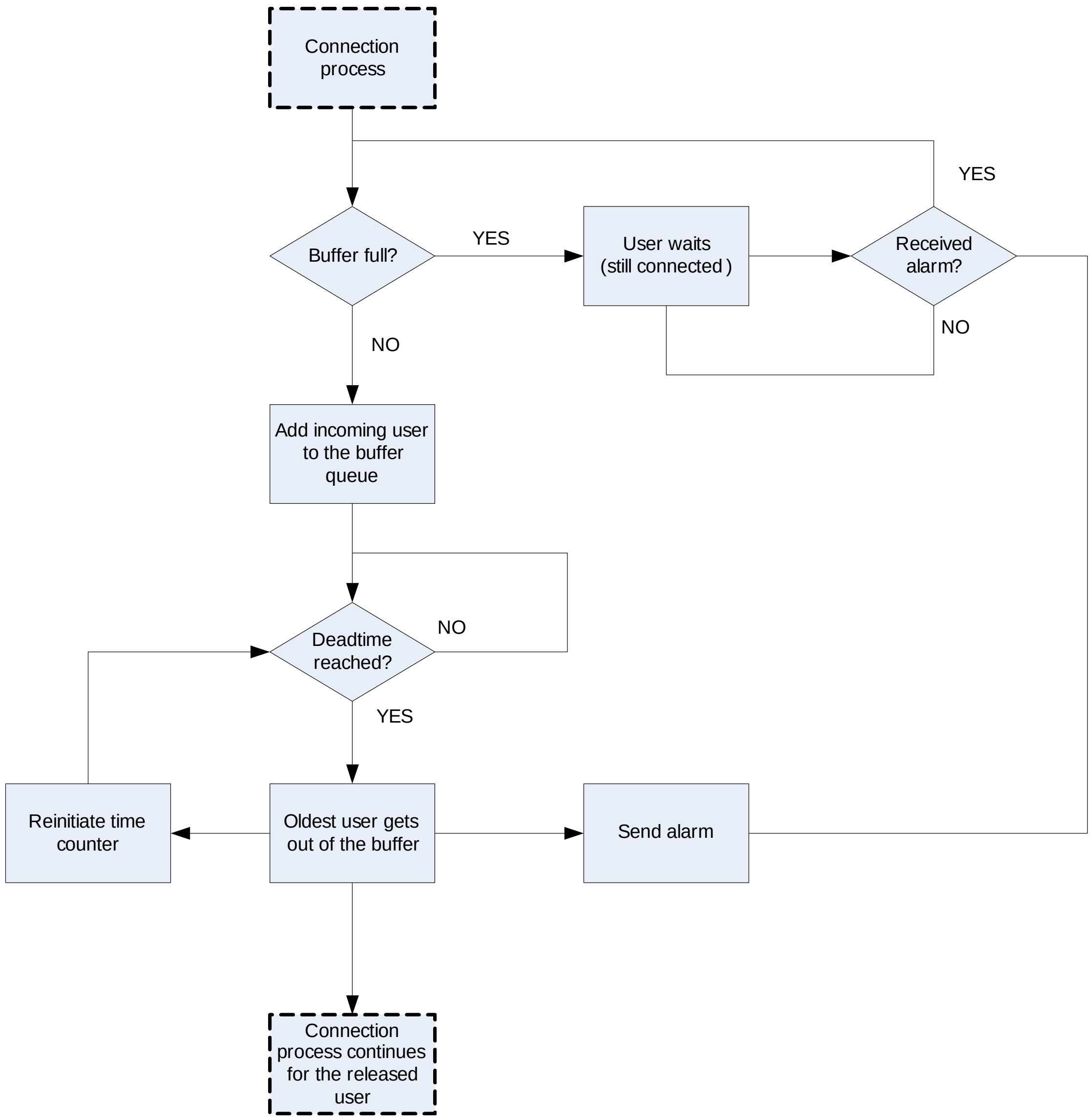

Figure 2-11: Connection scheduling process.

The admission control scheduling process sorts out the users to grant them access to the mode, provided that it is available. In case the mode can not accept the user, there are three possible actions. The first one is to perform a handover for low-priority sessions, if it can release enough resources to accept the user. The second one is to degrade the QoS of the low-priority sessions, still to obtain resources for the incoming user. Finally, if the two previous actions can not be achieved and if none of the modes can accept the user, a handover to legacy RANs is considered, provided that the application can be executed in one of them.

We are aware of the fact that our first action, the handover of low-priority sessions in order to release resources, is costly, and not necessarily profitable. However, in a QoS approach, we rather try to handover low-priority users and guarantee them an acceptable QoS than immediately degrade their bit rate.

If the user can not be connected to a WINNER mode (in case of buffer full or of no available mode), the application compatibility with a legacy RAN is verified. If none of the legacy RANs can provide the application, the user is rejected; otherwise the availability of the compatible RANs is checked. Similarly to before, if the network is not available, handover and then QoS degradation of low-priority sessions are tried to release as much resources as needed. In case it works, the user is finally accepted and low-priority users are redirected or see their bit rate decrease; otherwise the user is rejected.

The user's application request is made at the Mobile Terminal. The ACS determines the default mode corresponding to the application using the Service Set list, and its availability is checked at the SRRMW once the user has been through the scheduling process. If the mode is not available, the CoopRRM checks if an ISHO for low-priority sessions is possible, in order to free resources for the incoming user. This verification takes place at the CoopRRM since the availability of the RANs that would accept the low-priority users and their compatibility with the users' applications have to be checked. The user is then accepted. If no ISHO can be performed, the SRRMW checks if low-priority sessions can have their bit rate decreased in order to release resources. The system ensures that low-priority sessions QoS will remain acceptable, but sufficient resources can maybe be withdrawn. The user is then accepted.

When a user requests a new application, the default mode corresponding to the application is determined, and the user is placed in the buffer queue if the latter is not full. If the buffer is full, the mode is considered not available, and the user is sent to another mode or RAN, with respect to their compatibility with the application. If the buffer has still room, the user is placed in a certain position in the queue, according to his global priority. To do this, the entering user's priority is compared to the global priority of each user in the buffer, in increasing order. If the entering user's priority is greater than or equal to the current user's priority, the following case is checked, until the entering user's priority is strictly smaller than the current user's one. When this point is reached, the entering user takes the place of the other user in the queue, and every user is moved to the following rank after that case.

Similarly to the connection scheduling process, a time counter determines when the first-positioned user gets out of the queue. This happens after a delay that still has to be determined.

Let us take an example to illustrate this process. User U5, which basic user's priority is 3, requests to use Application 8. According to Table 2-9 and Table 2-10, user U5_8's global priority is 30+5 = 35. The buffer is the following when U5_8 enters it:

Buffer's rank	1	2	3	4	5	6
User's ID	U3_2	U2_12	U1_5	U4_3		
Global Priority	11	34	52	63		

Table 2-9

User ID	**User's basic priority**	**Application priority**
U1_5	5	2
U2_12	3	4
U3_2	1	1
U4_3	6	3

Table 2-10

Buffer's rank	1	2	3	4	5	6
User's ID	U3_2	U2_12	U1_5	U4_3		
Global Priority	11	34	52	63		
U5_8 global priority	35	35	35			
Comparison	35>11	35>34	**35<52**			

Table 2-11

Since U5_8 global priority is smaller than U1_5 global priority, U5_8 will take the place of U1_5 in the buffer, and U1_5 and U4_3 will have a rank decreased of one position.

Buffer’s rank	1	2	**3**	4	5	6
User’s ID	U3_2	U2_12	**U5_8**	U1_5	U4_3	
Global Priority	11	34	**35**	52	63	

Table 2-12

If it is still impossible and when all modes have been tried, the CoopRRM checks if there are compatible legacy RANs that could provide the application (see Chapter 3). If it is the case, the availability of the RANs is checked at the SRRML; otherwise the user is rejected by the CoopRRM. If the RAN is available, an ISHO is performed for the user. On the contrary, if the RAN is not available, the same verifications as before are made for the legacy RAN: a possible ISHO for low-priority users at the CoopRRM level, and ultimately the low-priority sessions QoS degraded at the SRRML level so that the user can be accepted.

The aim of the admission control scheduling algorithm is to organise the users so that they can not start to use their application(s) at the same time and to prioritise them, depending on their user’s and application’s priority.

Therefore, users having higher priority are always served before the others, and depending on the degree of interactivity the application requires, they will also be first served. Implementation of a When a user requests a new application, the default mode corresponding to the application is determined, and the user is placed in the buffer queue if the latter is not full. If the buffer is full, the mode is considered not available, and the user is sent to another mode or RAN, with respect to their compatibility with the application. If the buffer has still room, the user is placed in a certain position in the queue, according to his global priority. To do this, the entering user’s priority is compared to the global priority of each user in the buffer, in increasing order. If the entering user’s priority is greater than or equal to the current user’s priority, the following case is checked, until the entering user’s priority is strictly smaller than the current user’s one. When this point is reached, the entering user takes the place of the other user in the queue, and every user is moved to the following rank after that case.

2.2.3.5 Disconnection process

The general flow chart fort he disconnection process is shown in Figure 2-12.

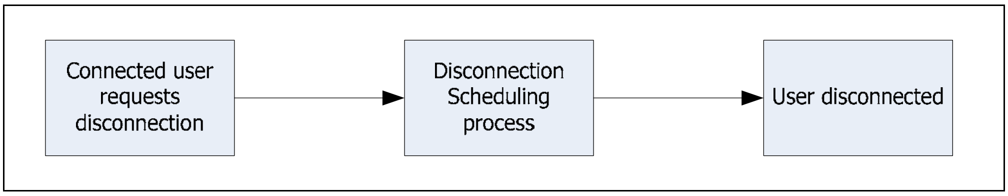

Figure 2-12: Disconnection process

When a user wants to be disconnected, he will first go through the disconnection process as shown Figure 2-12. Users however, again go through a scheduling process during a disconnection. This is shown in Figure 2-13.

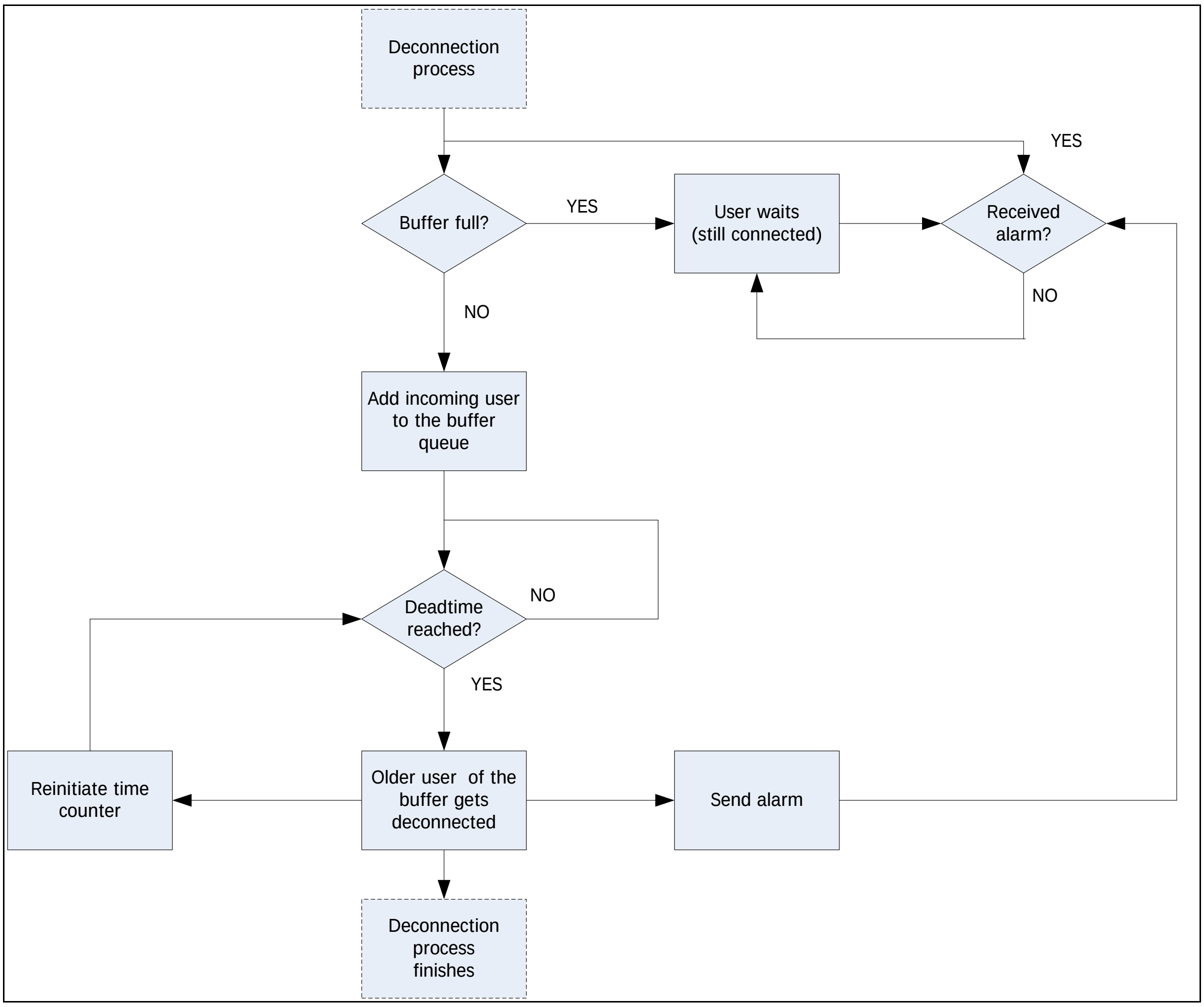

Figure 2-13: Disconnection scheduling process

The disconnection scheduling process is works in pretty much the same way as the connection scheduling process: If the buffer is full, the user is placed in a waiting state, until an alarm corresponding to the release of a place in the buffer is triggered, indicating that a waiting user can enter the queue. The users are ordered in the queue following a “First In First Out” rule. The first-ranked user gets out after the time counter has reached the dead time. The time counter is then reinitiated, the alarm sent to waiting users, and each user in the queue increase his rank from one position, i.e. user 2 takes the empty place of just left user 1, user 3 takes the place of user 2, etc

2.2.3.6 Expected download time

The **expected download time** is expressed in seconds and depends on the type of application used and on the size of the file the user wants to transfer or receive. This information will be computed at the ACS and added to the application profile.

The ACS will get the file size information either at the MT, if the user wants to upload data; or at the IP backbone corresponding interface if the user wants to upload an Internet webpage for example. This file size is then divided by the throughput value computed for the mode which the user should be assigned to. Therefore, when a user requests a file exchange application, the expected download time is computed and the mode which is most capable of satisfying the user from an application transfer success point of view is known.

The throughput value is used in the formula, because we assume that the network status will not change roughly between the time the user requests the connection and the time the download time is computed. Thus, the prediction concerning the mode which should keep the user the longest is fair.

$$Expected\ upload\ /\ download\ time = \frac{File\ size}{Throughput}$$

2.2.3.7 Congestion Thresholds

Congestion is a phenomenon occurring in communication systems of variable geometry, or systems in which the number of users is dynamically changing. In these communication systems, the traffic characteristics can vary over time, causing non deterministic utilisation of the system and allowing congestion situations. Indeed, congestion happens if the total amount of traffic entering the system within a predefined time interval is greater than the outgoing capacity of the system in the same time interval. In [Mon96], a communication system is defined congested whenever the functioning of communications services is affected in a way that is perceptible to their users.

Congestion can be characterised as the situation in which the load is too high for the network to be able to act normally. In [Ber01], congestion is said to occur if the load reaches 70% of the theoretical maximum capacity. Therefore, with the load that we defined previously, our network will get congested when the allocated bandwidth, in the data rate sense, will represent 70 % of the total bandwidth of the network.

When thinking about the practical implementation of a congestion control algorithm, it is obvious that two different load values are required to identify when the algorithm has to enter the congestion resolution phase and when the algorithm can stop the congestion resolution phase. If the same threshold value was used for both entrance and exit, the load value after congestion resolution would be very close to the threshold value which triggered the congestion control process. Therefore, a little change in the activity of the network would trigger the whole algorithm once again. The situation of the network would not be stable enough after congestion resolution.

The highest threshold value will be used to enter the congestion resolution process while the lowest one will be used to exit it. The lower value is thus triggering the congestion recovery phase, where the users whose QoS was degraded can recover it. However, this lower value cannot be too low; otherwise the congestion resolution process might reject or drop more users than necessary. The two threshold values will be defined around the 70% figure previously introduced, and we will test several values for the simulations.

In order to avoid too prompt triggering of the congestion resolution algorithm, a time step is defined that corresponds to a hysteresis. The algorithm will be triggered if the load value has remained higher than the triggering threshold value during a certain time, corresponding to the time step. This time step is used to make sure that the algorithm is only used when it is indispensable and not in shallow congestion situations. This is a relevant process because it avoids as much as possible to block the network and it allows the network to operate more often in the 70% zone in which it is optimally utilised (maximum resources allocation with no congestion).

2.2.3.8 Traffic modeling

Important application’s parameters for the traffic modeling and simulation are the duration and the data rate. Regarding the duration, we have applied what was described earlier, or every application runs for its expected download time. Concerning the data rate, we have chosen to use for the simulated data rate the average of the minimum and maximum required data rates expressed in Table 2-4. The simulated data rate values are shown in Table 2-13.

For the simulation the modeling of the users has been done through the use of the user profile. The choice has been made to allow the user to run only one application at a time, which is not a critical assumption since, from a simulation point of view, two users with the same priority running respectively application *x* and *y* are the same than one user running both applications x and y.

At the simulation level the important parameters for a user are the location, the priority level and the application. For simulation handling matters we have defined some groups of users, which contain users having all the same location, same priority and same application. We have decided not to specify the location of the groups of users. The groups are then randomly placed within the area of study following a uniform distribution. On the other hand, the priority level of the group is to be manually assigned. The user is then associated with an application, that is to say with a certain running time, the application’s expected download time, and a certain data rate, the application’s simulated data rate, as previously described.

The application repartition within the users is of course meant to fit the penetration factors previously introduced and this is manually ensured when launching the simulations. The users’ repartition is chosen for the simulation with respect to the user priority. The results are shown in Table 2-14 and Table 2-15.

Service Class	Minimum required data rate (kb/s)	Maximum required data rate (kb/s)	Simulated data rate (kb/s)
1. Real Time Collaboration and gaming	1x1024	20x1024	10x1024
2. Geographic real time datacast	2x1024	5x1024	3.75x1024
3. Short Control messages and signalling	8	64	28
4. Simple interactive applications	64	512	224
5. Interactive high multimedia	2x1024	5x1024	3.75x1024
6. Geographic interactive multimedia broadcast	2x1024	5x1024	3.75x1024
7. Interactive ultra high multimedia	1x1024	50x1024	25x1024
8. Simple telephony and messaging	8	64	28
9. Data and media telephony	64	512	224
10. Geographic datacast	64	512	224
11. Rich data and media telephony	2x1024	5x1024	3.75x1024
12. LAN access and file services	512	50x1024	24.75x1024
13. Multimedia messaging	8	64	28
14. Lightweight browsing	64	512	224
15. File exchange		5x1024	5x1024
16. Video streaming		5x1024	5x1024
17. High quality video streaming	2x1024	30x1024	14x1024
18. Large files exchange	1x1024	50x1024	25x1024

Table 2-13: Simulated Data Rates

User priority	User profile	Repartition
1	UP1	16.8 %
2	UP2	60.7 %
3	UP3	22.5 %

Table 2-14: User's Repartition

Simulation duration (s)	Group size	User ID in the group	User entrance in the simulation (s)
600	100	1	0
600	100	2	6
600	100	3	12
600	100	4	18
…	…	…	…

Table 2-15: User's entrance calculation

2.3 Cooperation between WINNER and the legacies RANs

The WINNER system will be a flexible system that will be able to handle the provision of future mobile multimedia services offering an optimization of capacity in the air interface by means of efficient RRM algorithms. These functionalities are very important in the framework of future systems because the system relies on them to guarantee a certain target QoS, to offer high capacity and so on. However, the WINNER system cannot be considered as a stand-alone wireless system, it will have to co-exist with a number of legacy wireless systems. WINNER will be able to cooperate with these legacy systems in order to maximize the efficiency of the system, to serve the users with best QoS and to make easily the migration from the existing systems to the WINNER system. In WINNER I WP4 the work was focused on making the WINNER system cooperate with the legacy networks. There were defined many RRM mechanisms to ensure and actualize this cooperation. These algorithms include mobility management (inter-system handover), admission and congestion control and QoS management for heterogeneous systems. In this document it is presented also a fuzzy-logic based handover algorithm for the cooperation between WINNER and legacy systems.

2.3.1 Mobility management between WINNER and the legacy RANs

2.3.1.1 System Common Control Support

The network is able to interact with the legacy RANs through generic 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.

To enable fast frequency scanning and cell registrations to the proper RATs for the Multi-mode UTs, the WINNER system broadcasts 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.

A step further is the interworking between WINNER system and a global spectrum coordination system [Cor05][Moh05] when applicable. The global spectrum coordination system operates on a set of out-band signalling carrying the spectrum coordination w.r.t. the operators. Such interworking enables the advanced spectrum management by WINNER system.

To enable fast frequency scanning and cell registrations to the proper RATs for the Multi-mode UTs, the WINNER system should be able to interwork with the legacy systems in order to allow the legacy system broadcast the WINNER mode and carrier frequency in the evolved broadcasting message structure, for instance the SIB (System Information Block) of WCDMA system. Such system information can be the frequency allocation to the specific legacy RATs, and is determined by the operator's policy on inter-RAT interworking.

2.3.1.2 Enabling Functions for Intersystem Mobility Management

It is assumed that when a UT camped in WINNER, it could handover to a legacy RAN when it lost the coverage of WINNER system. This situation is relevant especially in the initial deployment with very limited number of WINNER BS, congestion in WINNER cells.
Hereafter, is presented the handover process and signalling of a UT from WINNER to a legacy RAN, as exemplary network have been chosen UMTS. In this process the SRRMW, CoopRRM and SRRM exchange information and message, as reference is section devoted to architecture are described this logical entities:

- SRRMW (associated to the WINNER ACS) detects a situation in which the UT service requirements can not be satisfied in any WINNER RAN mode (e.g. lost of coverage o QoS requirements degradation)
- SRRMW sends an indication message with measurements and statistics to the CoopRRM along with an intersystem handover request.
- CoopRRM uses periodic or ad-hoc measurements on the legacy RAN candidates to take a decision on handover. In case of ad-hoc measurements, the CoopRRM sends a measurement request to the SRRML of target RAN, and this entity answer with a measurement report.
- If CoopRRM decides to handover to UMTS, CoopRRM sends a HO_request message to the UMTS SRRML, indicating the identity of the mobile terminal and other useful information (on its QoS requirements, for instance)
- SRRML sends a Hard_HO_request message to L3 of the UMTS Core Network.
- L3 of the Core Network sends a Hard_HO request message to L3 of UTRAN, that starts the transition from idle mode to CELL_DCH state for that mobile terminal.

Consequently, this requires defining a new message between SRRM and L3 of the CN (or of the UTRAN), and also defining a new message between CoopRRM and SRRML.

Concerning measurements/information gathering, each RNC will have to send periodically or on request a message containing useful information to SRRM, which will then have to send another message with this information to CoopRRM. So the efficiency and the delay will depend on the location of SRRM. The closer it is from the RNC, the better (less delay). We could imagine that the RNC sends information to SRRM periodically without any filtering, and then that SRRM only sends information to CoopRRM on request. This would decrease the signalling load between SRRM and CoopRRM (which may be problematic if both entities are too distant). Figure 2-14 depicts a simplified the inter-RAN handover process

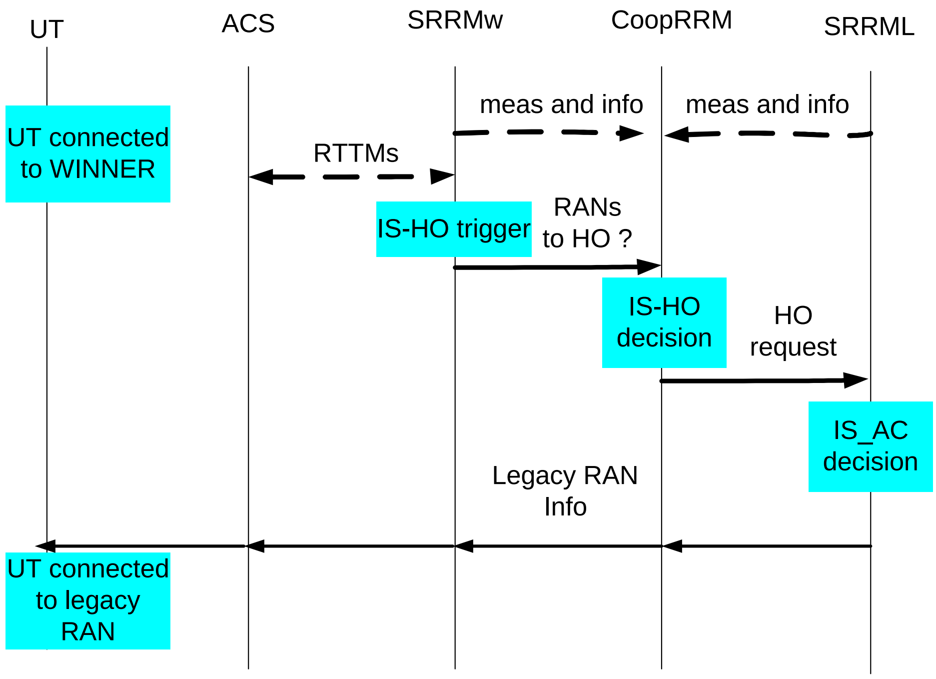

Figure 2-14: Handover process from WINNER to a legacy RAN

2.3.2 Fuzzy logic based mobility management

2.3.2.1 Introduction

In this section, the focus is on the mobility management between WINNER and legacy RANs. Hereafter, are defined advanced handover algorithms based on techniques such as fuzzy logic and automatic learning. Fuzzy logic is a simple and fast solution to provide a conclusion from imprecise, noisy or incomplete inputs. Fuzzy logic is based on simple "IF X AND Y THEN Z" rules rather than complicated mathematical models.
In this section, a scheme of handover between WINNER and legacy RANs, based on fuzzy logic, is described.
In this document, the first part deals with fuzzy logic: motivations and mathematical basics. The second one presents a scheme of fuzzy logic based handover between WINNER and legacy RANs.

2.3.2.2 Fuzzy logic overview

2.3.2.2.1 Fuzzy logic use motivations

Fuzzy logic is a robust solution in presence of imprecise or incomplete and noisy inputs as well as measurement tools imprecision. Fuzzy Logic can control systems that could not be modelled mathematically. The fuzzy system uses a list of rules to control the system. System behaviour can be tuned, simply, by modifying the appropriate rules.
In one hand, fuzzy logic is adapted to radio resources management because of radio environment fluctuations and uncertainty (measurements averaging, shadowing, traffic model …). In the other hand, thanks to fuzzification, it is possible to compare quantities from heterogeneous RANs. In complex systems, fuzzy models, based on simple IF-THEN rules, gives more easily assimilated information than precise models. Nevertheless, rules definition requires a good knowledge of systems and prior field experience. Learning techniques improve the fuzzy system parameters and enhances its performances accordingly.

2.3.2.2.2 Fuzzy logic bases

2.3.2.2.2.1 <u>Membership functions</u>

Unlike the classical logic, there are not absolute (0 or 1) membership degrees but relative membership degrees between 0 and 1. Membership degree is the possibility that an element belongs to the fuzzy set. Membership functions are defined thanks to a prior experience in a first time and then optimized, by means of learning, according to system response.
A is a fuzzy set on X, it is noted as [DUB80]:

- If X is finite: $A = \sum_{i=1}^{n} \mu_A(x_i) / x_i$

- If X is not finite: $A = \int_X \mu_A(x)/x$

2.3.2.2.2.2 Basic operations on fuzzy sets [DUB80]

$$\mu_{A\cup B}(x) = \max(\mu_A(x), \mu_B(x)), \forall x \in X$$

$$\mu_{A\cap B}(x) = \min(\mu_A(x), \mu_B(x)), \forall x \in X$$

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x), \forall x \in X$$

2.3.2.2.2.3 Fuzzy control

The control problem [GLO99] is to find inputs that reach a defined goal under certain constraints on the control sequence and on the intermediary states of the system.
The rule "If X is A_i, then (if Y is B_i, then Z is C_i)" is translated using the "min" operator.

$$\begin{aligned}\mu_{R_i}(X,Y,Z) &= \min(\mu_{A_i}(X), \min(\mu_{B_i}(Y), \mu_{C_i}(Z))) \\ &= \min(\mu_{A_i}(X), \mu_{B_i}(Y), \mu_{C_i}(Z))\end{aligned}$$

In the case where there are n fuzzy rules, the resulting rule is:

$$\mu_R(X,Y,Z) = \max_i \mu_{R_i}(X,Y,Z)$$

There are two frequently used types of Fuzzy Inference System (FIS): Mamdani and Takagi-Sugeno.

2.3.2.2.2.3.1 Mamdani method

The fuzzy inference in Mamdani method is made up of the following steps for the input vector ($x_1,\ldots, x_n$):
Compute the membership degree to each fuzzy subset of each input.
Determine the truth value of each rule:

$$\alpha_i(x) = \min_j (\mu^i_{A_j}(x_j)),\ j = 1 \ \ to \ \ n$$

Compute the contribution of each rule

$$\mu_i(y) = \min(\alpha_i(x), \mu_{B^i}(y))$$

Rules aggregate:

$$\mu_i(y) = \max_i(\mu_i(y))$$

To obtain a precise conclusion, the result must be defuzzified using the gravity centre method

2.3.2.2.2.3.2 Takagi-Sugeno method

Takagi-Sugeno FIS conclusion is function of the inputs.
For the input vector ($x_1,\ldots, x_n$), these steps are executed:
Compute the membership degree to each fuzzy subset of each input.
Determine the truth value of each rule:

$$\alpha_i(x) = AND(\mu^i_{A_1}(x_1),\ldots,\mu^i_{A_n}(x_n)),\ j = 1 \ \ to \ \ n$$

Where AND is the conjunction operator.
The output of the FIS:

$$y = \frac{\sum_{i=1}^{N} \alpha_i(x) \times s_i}{\sum_{j=1}^{N} \alpha_j(x)}$$

The Takagi-Sugeno method can be considered as a particular case of Mamdani method when the fuzzy output subsystem is a single point.

2.3.2.3 Mobility management scheme using fuzzy logic

2.3.2.3.1 Introduction

This part aims at defining an intersystem RRM scheme using fuzzy logic. First, the handover algorithm inputs are defined. The handover decision criteria could be network criteria or user criteria and are system dependant. We determine handover criteria and specify them per RAN (WINNER, UMTS and WLAN).
Then, the fuzzy block components are described thoroughly: the fuzzification function that converts an input value into linguistic variables by means of membership functions, the inference function that apply the fuzzy rules and the decision choice function. Finally, the handover model using fuzzy logic is defined.

2.3.2.3.2 Handover decision criteria

2.3.2.3.2.1 Measurements per RAN

In the table below, we summarize the corresponding measurements to each system.

Measurements per System	**UMTS**	**WLAN**
Signal strength	CPICH RSCP, CPICH Ec/N0	RSSI
Signal Quality	BLER	BER
Load	- P_dl/P_dl_max - Number of channel elements - Number of available codes	Relative occupied bandwidth

Table 2-16: Measurements list per system

In UMTS FDD mode, the load can be evaluated through the total transmitted power in downlink: P_dl/P_dl_max.

2.3.2.3.2.2 User terminal parameters

User priority
Service class
QoS requirements
User velocity

- Location information

Terminal capabilities
Battery status

- Price

2.3.2.3.3 Description of the fuzzy block

The fuzzy logic block is made up of:

- Fuzzification function that transforms the numerical inputs into fuzzy variables. Measurements are detailed and membership function per measurement determined.
- Decision rules application function: the defined inter system handover rules are applied to the fuzzy inputs. The output is a set of triplets made up of an inter system handover decision and an evaluation degree.
- Decisions evaluation function: choice of the handover decision and the suitable target system.

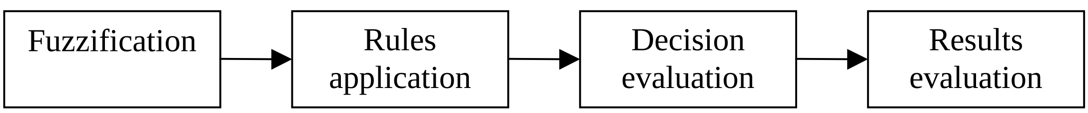

Figure 2-15: Fuzzy system components

2.3.2.3.3.1 Fuzzification

Thanks to the *fuzzification function*, the inputs are transformed into linguistic variables coupled with a membership degree. For each variable input type, fuzzy sets (LOW, HIGH for example) are defined. Every fuzzy set is characterized by a membership function. A numerical input is projected on the membership function of each fuzzy set and in output, we have a set of n couples (fuzzy set i, degree of membership to fuzzy set i). n is the number of fuzzy sets composing the variable. The result is assigned to a vector of membership degrees to the different fuzzy sets with a dimension equal to the number of fuzzy sets relative to the variable.
For instance, if the input is Signal Strength (SS), it can be classed as high, medium or low as shown in the graph above.
Let's consider the variable X, let A_i i within {1,.., n} be the fuzzy set relative to X. The fuzzification gives the vector $(\mu_{A_1}(x),..., \mu_{A_n}(x))$.

2.3.2.3.3.2 Fuzzy rules application

The *fuzzy rules application function* applies a number of rules to the outputs resulting of the fuzzification operation. Rules, with an IF-THEN structure, decide whether or not the mobile should hand over to another system and choose the target entity in the affirmative case.
An example of a simple fuzzy rule could be:

IF *(Current system = WINNER_FDD)* ***AND*** *(Mobile terminal velocity = LOW)* ***AND*** *(WLAN coverage = MEDIUM* ***OR*** *HIGH) AND (WLAN load = LOW* ***OR*** *MEDIUM)* ***THEN*** *(handover to WLAN)*

This rule deals with situations where the UT is attached to WINNER_FDD and it is moving with a low velocity. If WLAN offers a satisfactory coverage and load levels, then a handover to WLAN is suggested.

The output of the inference step is a set of handover decisions specifying the target RAN with membership degrees.

2.3.2.3.3.3 Decisions evaluation function

In this final step of decision, each decision is evaluated thanks to a score function. This score function is defined according to rules priority. The decision with the highest score will be chosen.

2.3.2.3.4 The fuzzy logic based handover scheme

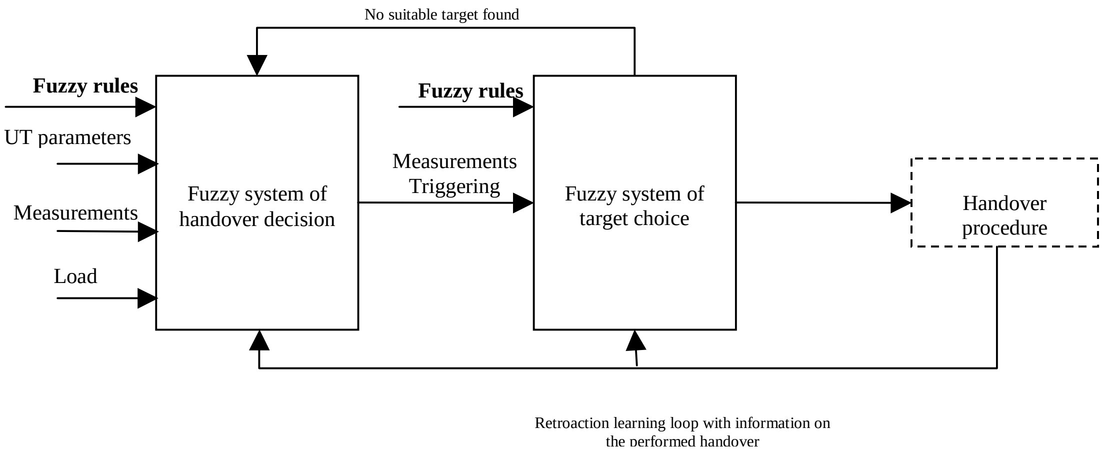

Figure 2-16: Fuzzy logic based handover scheme

Fuzzy system of handover decision, located in CoopRRM, receives information such as UT parameters, measurements and load. According to this information, it decides to trigger handover measurements on the other present modes or RANs. Then, it decides to perform the handover to the suitable mode or RAN. Finally, The SRRM of the new serving RAN must report the results of this handover to the CoopRRM. This Retroaction loop with information on the performed handover is used by the learning process. The fuzzy system learns from the results of the previous decisions how to enhance the rules parameters. Obviously, a first phase of offline learning is needed, using simulations and traffic prediction. Afterwards, the optimized fuzzy RRM system is actually applied; it could also be enhanced by means of online learning, to fit further traffic variations.

2.3.3 Congestion Avoidance Control

As it was analysed in section 2.2.2 the Congestion Avoidance Control is a very important mechanism for Wireless Networks and it includes the Admission Control mechanism and the Load Control mechanism. For the cooperation between WINNER and legacy RANs these two mechanisms were extensively researched in WINNER I and algorithms were defined and tested.

2.3.3.1 Admission Control

In WINNER I it was defined in details an admission control algorithm for the cooperation between WINNER and legacy system. The detailed description can be found in [WIND42][WIND43]. The algorithm’s most important criterion is the load of the networks. The formulas for the computation of the load of the legacy system were also defined in details. The algorithm’s efficiency and functionality was evaluated and assessed by system level simulations, using an emulation of the WINNER system via high-datarate WLAN network with micro- and macro-cells for the short area and wide area WINNER cells. The simulation results proved that the algorithm is very efficient and supports the cooperation between WINNER and the legacy systems. The simulation results can be found in [WIND44][WIND45].

2.3.3.2 Load Control

In WINNER I it was defined in details a reactive load control algorithm for the cooperation between WINNER and legacy system. The detailed description can be found in [WIND42][WIND43]. The algorithm’s actions on resolving congestion situations are:

- interaction with admission control to reject new user requests
- force ongoing users that have high load to handover to other networks
- decrease the load of ongoing users that have high load
- drop ongoing users with high load

The algorithm’s efficiency and functionality was evaluated and assessed by system level simulations, using an emulation of the WINNER system via high-datarate WLAN network with micro- and macro-cells for the short area and wide area WINNER cells. The simulation results proved that the algorithm is very efficient and supports the cooperation between WINNER and the legacy systems. The results showed that the algorithm performs very well even in situation of very high incoming traffic and it keeps the network decongested in any case. The simulation results can be found in [WIND44]]

2.3.4 QoS based management

The general flow chart of the QoS management process is shown in Figure 2-17

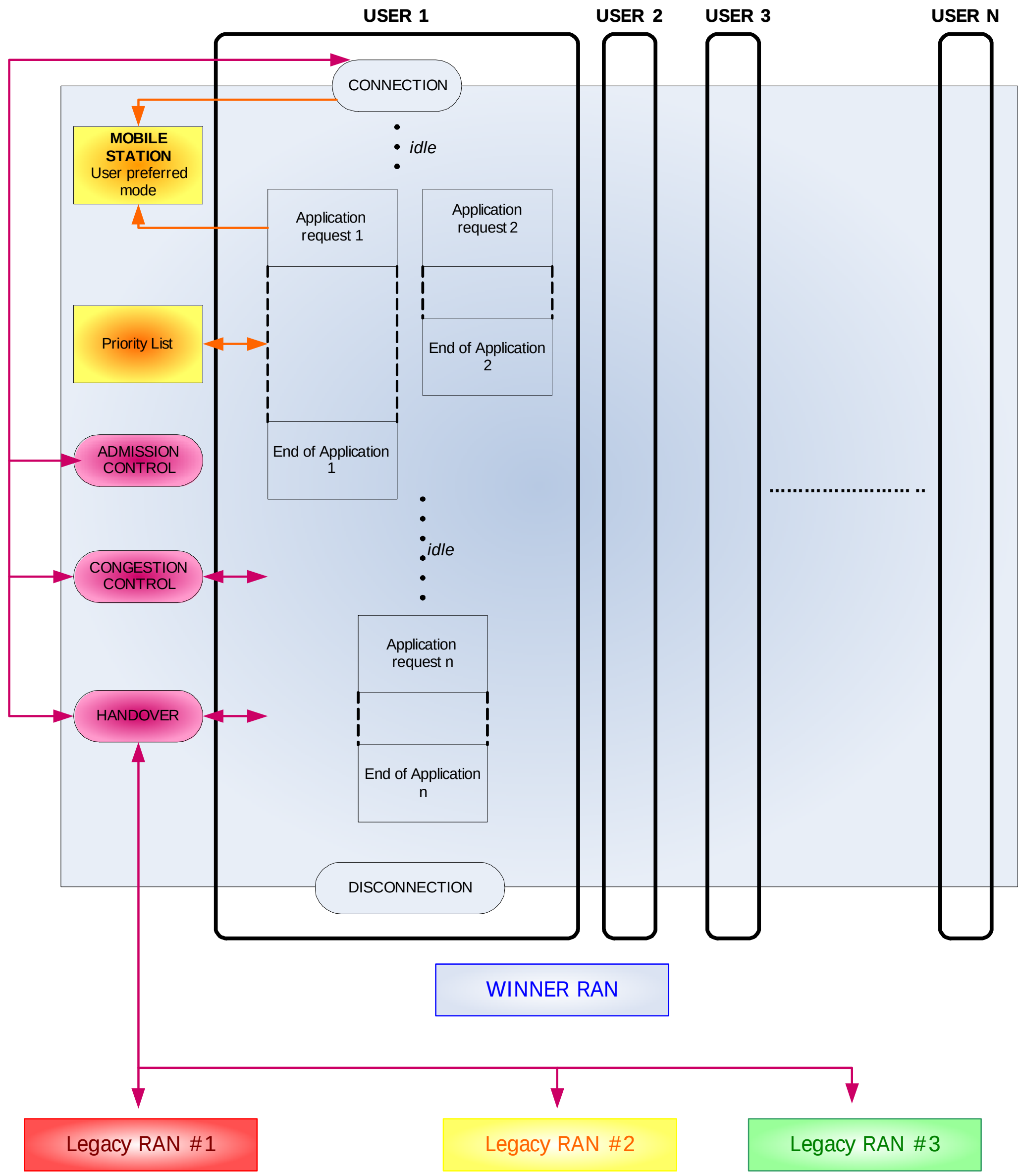

Figure 2-17: General flow chart of the QoS management process

The connection process is the first step in the QoS approach. If the user does not meet the conditions necessary for a connection, or if the network is not available, the user will be rejected. On the contrary, if the user is to be accepted, he will be directed as much as possible to the network that suits the best his future requirements. However, it is unlikely that the user gets rejected in the connection phase, unless he has no rights to be granted access to it, or if the network is congested. Figure 2-18 represents the connection algorithm. We assume that the user can request a connection and remain idle without using any application during a certain time. When a user wants to connect to the WINNER network, it is checked that he belongs to the network operator. In the contrary case, if agreements between the home operator and user's operator exist, the user will be accepted thanks to an inter-operator HO, otherwise he will be rejected. If the user belongs to the operator, verification is made in order to know if the user has access to the WINNER RAN.

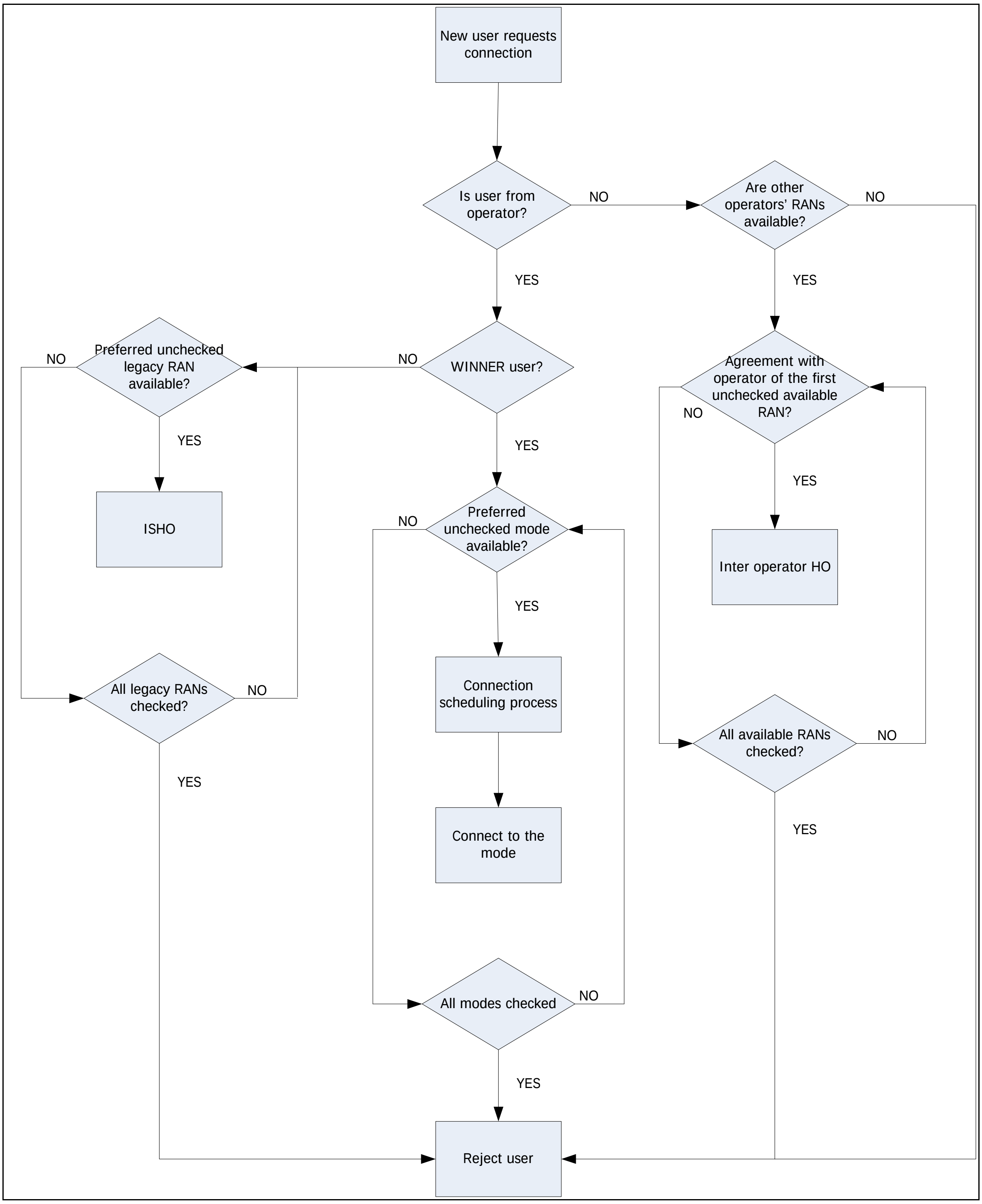

Figure 2-18: General flow chart of the connection process

2.3.4.1 Key Performance Indicators for QoS based management

Key Performance Indicators (KPIs) are a set of measurements used to keep track of a network status over the time. KPIs can be split in two types whether they describe the network's resources or the QoS delivered. The main KPIs related to QoS can be measured in any type of packet-switched network. They are listed and defined below:

- **Delay** (expressed in seconds)

Delay, (which can be referred to as latency in the case of WLAN or a system operating in a short-range mode), expresses the time needed for one packet of data to get from one designed point to another. The round-trip delay is measured by the time taken for sending a packet that is returned to the sender. From this, the one-way delay can be calculated, being half of the round-trip delay. A delay much longer than expected indicates congestion in the network.

- **Jitter** (expressed in seconds)

Jitter represents the delay variation of the received packets over time. Indeed, packets sent at a constant rate are not necessary received at a constant rate, due to network behaviour–especially congestion. Jitter is hence the measure in time of the irregularity of the packets transmission of the. Jitter effects are cancelled through the use of a buffer at the receiving end.

Several formulas for jitter calculation can be defined. Jitter can first be calculated as a raw spreading of the delay around the expected delay:

$$Jitter = \frac{1}{N}\sum_{n=1}^{N}(\tau_n - \tau_0)$$

where N is the number of transmitted packets;
τ_n is the delay in seconds of the nth received packet;
τ_0 is the expected delay in seconds.

If no expected delay is available, another reference must be chosen, for example the delay of the first received packet. The formula is hence:

$$Jitter = \frac{1}{N-1}\sum_{n=2}^{N}(\tau_n - \tau_1)$$

where N is the number of transmitted packets;
τ_n is the delay in seconds of the nth received packet;
τ_1 is the delay in seconds of the first received packet.

Jitter can also be calculated with reference to a mean delay of the previously received packet, using a recursive formula:

$$\forall n \in [\![1, N]\!], Jitter(n) = \tau_n - \frac{1}{n-1}\sum_{i=1}^{n-1}\tau_i$$

where N is the number of transmitted packets;
τ_n is the delay in seconds of the nth received packet;
τ_i is the delay in seconds of the ith received packet.

This formula has the advantage of providing a value of jitter for each received packet, without having to wait for the end of the transmission of the group of packets. This can be interesting since we need to have the information at each time.

Moreover, it may be important to calculate the absolute jitter. Indeed, with the previous raw jitter formulas, the delays of packets arriving late can be minimized by the delays of packets arriving early, and information might be hidden. That is why we also define an absolute jitter, which will have in any case a value equal or greater than raw jitter. The formulas are listed below, respectively for jitter with reference to an expected delay, jitter with reference to the first arriving packet delay and jitter with reference to mean delay:

$$Jitter = \frac{1}{N}\sum_{n=1}^{N}\left|\tau_n - \tau_0\right|;$$

$$Jitter = \frac{1}{N-1}\sum_{n=2}^{N}\left|\tau_n - \tau_1\right|;$$

$$\forall n \in [\![1, N]\!], Jitter(n) = \left|\tau_n - \frac{1}{n-1}\sum_{i=1}^{n-1}\tau_i\right|$$

- **Peak user data throughput** (expressed in bps)

Peak user data throughput is the measure of the maximum rate achieved during the transmission of data in the network. This KPI must refer to a single user.
For this measurement, an instantaneous user data throughput must be available from the network, i.e. an instantaneous value of the datarate for each user.

The formula for the peak user data throughput calculation can be:

$$PUDT = \max(IUDT(t))$$

where IUDT(t) is the instantaneous user data throughput function, in bps.

- **Mean user data throughput (expressed in bps)**

Mean user data throughput is the measure of the average rate achieved during the transmission of data in the network. This KPI must also refer to a single user.
The calculation is usually made by comparing the size of the transmitted data with the time of transmission of these data, both for uplink and downlink. The calculation may also be done with an integration of the previously mentioned instantaneous user data throughput function.

Uplink:

$$MUDT_{UL} = \frac{uploaded\ data\ payload}{upload\ time\ for\ data\ transfer}$$

Downlink:

$$MUDT_{DL} = \frac{downloaded\ data\ payload}{download\ time\ for\ data\ transfer}$$

2.3.4.1.1 Modelling the delay

Here we present one simple example of how to model the delay for the purpose of assessing its dependence on the network load. This approach has been adopted here for the simulations. In a low network load situation, the delay value τ can be represented as a typical delay τ_{typ}. When the load increases and gets in the congestion zone, the delay value will augment very quickly. The formula considers the influence of a congestion threshold parameter (CT) that shows when the congestion zone will be reached. Once this critical value has been reached, the CoopRRM will receive a request for handling the arisen congestion situation and an algorithm will be activated. Assuming that no significant change in delay may occur before the 40% load value is reached and that the higher delay value (i.e., for the 100% load value) must remain coherent for the chosen scenario, the delay can be expressed by Equation (1):

$$\tau = \tau_{typ} + 120.e^{\frac{L-\gamma}{12}} \quad (1)$$

Where τ_{typ} is the typical delay value in ms;

L is the load value, measured in percentage, of the total capacity of the cell; is a parameter that depends on the chosen congestion threshold and is expressed in percentage of the total capacity as given by Equation (2).

$$\gamma = 40 + \frac{CT}{2} \quad (2)$$

where CT is the load value, expressed in percentage of the total capacity, chosen to identify a congestion situation. This is the upper congestion threshold. shows the delay variations in a network with 20 ms as typical delay value. It represents the packet delay in ms versus the load of the network in percentage of the total capacity, for different values of the congestion threshold.

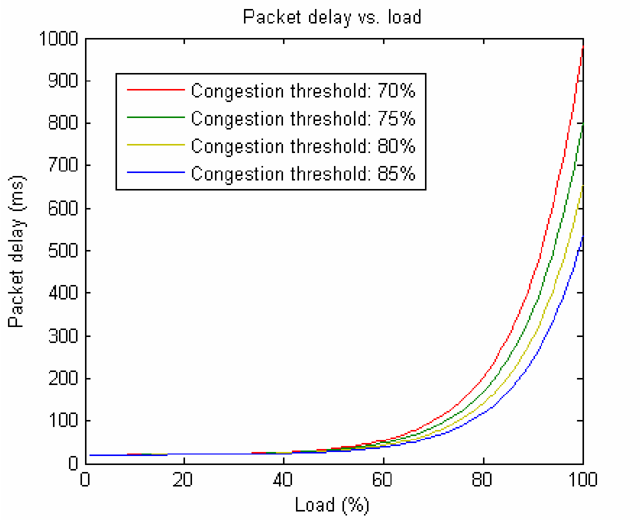

Figure 2-19: Delay modeling for different loads of the network [Mih06]

The delay is higher for higher congestion thresholds, which is a quality indicator for the QoS that a user will perceive when requesting a given application. The tolerable delay will depend on the application, for the identified for WINNER applications, all relevant parameters, including tolerable delays, download times, and required throughput, have been specified below.

2.3.4.1.2 Resources KPI's

Available Bandwidth and Throughput

Interesting KPIs concerning the network status are the **available bandwidth** and the **throughput**. In order to define these terms, hereafter we present some definitions. These definitions were given for a wired network context, but they are also valid in our case if considering the data traffic in particular. In this case, the word bandwidth does not represent a range of frequencies but is employed in a data rate sense. Indeed, a given frequency range can bear a corresponding data rate, depending on the coding scheme and multiplexing technique used, so the term is sort of misused:

- **Bandwidth**, in the data rate sense, is the speed at which a network element can forward traffic. It has two characteristics – physical and available, and both of them are independent of end hosts and protocol types;
- **Physical bandwidth** or **Capacity (C)** is the maximum number of bits per second a network element can transfer. The physical bandwidth of an end-to-end path is determined by the slowest network element along this path;
- **Utilisation (U)** is the percentage of capacity currently being consumed by aggregated traffic on a link or path;

$$U = \frac{Traffic}{C} \tag{3.1}$$

- **Available bandwidth (A)** is the capacity minus utilisation over a given time interval;

$$A(t_s, t_e) = Capacity - Traffic \tag{3.2}$$

$$\Leftrightarrow A(t_s, t_e) = C \times (1 - U)$$

Where: t_s is the time at which the measurement starts;

t_e is the time at which the measurement ends.

- **Throughput** is the amount of data that is successfully sent from one host to another via a network. It may be limited by every component along the path from source to destination host, including all hardware and software. Throughput also has two characteristics – achievable and maximum;
- **Achievable throughput** is the throughput between two end points under a completely given set of conditions, such as transmission protocol, end host hardware, operating system, tuning method and parameters, etc. This characteristic represents the performance an application in this specific setting might achieve.

Therefore, the available bandwidth is a measurement that indicates if there are still resources that users can exploit. The achievable throughput can be low even if there is still available bandwidth if for example a network element is limitative.

There are two ways of estimating the available bandwidth: the passive measurement, which consists in using the existing data transmission history, and the active probing, which consists in creating the situation in which it will be possible to measure the available bandwidth. The principle is that the sender sends a pair of packets echoed back by the receiver. By measuring the changes in the packet spacing, the sender can estimate the bandwidth properties of the path. presents several ways to estimate the available bandwidth, which is a difficult task, since it is a dynamic property depending on many factors.

Several methods exist such as Packet Bunch Mode (PBM), cprobe, Treno, Asymptotic Dispersion Rate (ADR). also presents Initial Gap Increasing (IGI) and Packet Transmission Rate (PTR) methods and the method of minimal backlogging. These methods vary parameters such as packet size, of the time interval at which packets are sent, etc. However, considering the duration of the project, we will not be able to implement these methods practically.

Network Load

The **load** of the network is a very important parameter that describes how much the network is utilised over time, in term of resources. Existing definitions for the calculation of the load of a network are very different depending on the type of network dealt with. In GPRS, it is only the ratio between available time slots and the total number of time slots. In WLAN, it is done either like in GPRS or with a ratio between collisions and transmissions. In UMTS, the definition is much more complicated as it embeds the influence of noise and of interferences, with different calculations for UL and DL.

In the WINNER case, the load computation method will be different for each mode of application, since each mode has a priori its own duplex technique. In order to simplify our work, we define load for WINNER only with respect to the bandwidth metric in the data rate sense, as it was defined previously. This way, we have a simple definition that can be used with all modes. Load then describes the utilisation of the network capacity, that is to say the quantity of occupied bandwidth.

With this assumption, the value of the network load is obtained by adding the value of bandwidth used by each user at a time, which is to say that we will use the previous definition of “Utilisation” to describe the load.

2.3.4.2 WINNER Users tracking process

In the architecture defined in Section 2, the RAN offering the best performances and then allowing the use applications in an optimal manner is obviously the WINNER RAN. Nevertheless, it can happen in case of WINNER overloading that a user owning a WINNER contract can not have access to WINNER at start-up. Following the admission control algorithm, he is then directed toward one of the legacy RAN. Even if this step can be satisfying at the beginning, it is important to plan to redirect him toward the WINNER network as soon as the latter becomes available again.

Hence we define a generic tracking algorithm to be placed on each of the legacy RANs. This algorithm is described in Figure 2-20

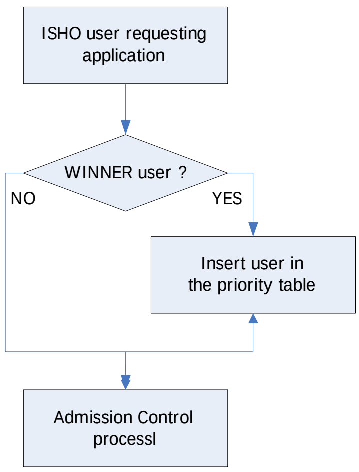

Figure 2-20: Tracking of the WINNER users

For the operation of this process, a small database has to be created. This database is placed at the CoopRRM. It contains all the exiled WINNER users sorted by priority with the same criteria as above, together with the RAN they are using. This database is updated each time a WINNER user arrives on a legacy RAN or ends an application that was running on a legacy RAN.

This information stored at the CoopRRM is useful for the operation of the repatriating algorithm. This is an ongoing algorithm running at the CoopRRM which regularly checks the status of the WINNER RAN. When it is available, it tries to redirect as much as possible exiled users on WINNER, beginning with high-priority users. This is shown in Figure 2-21.

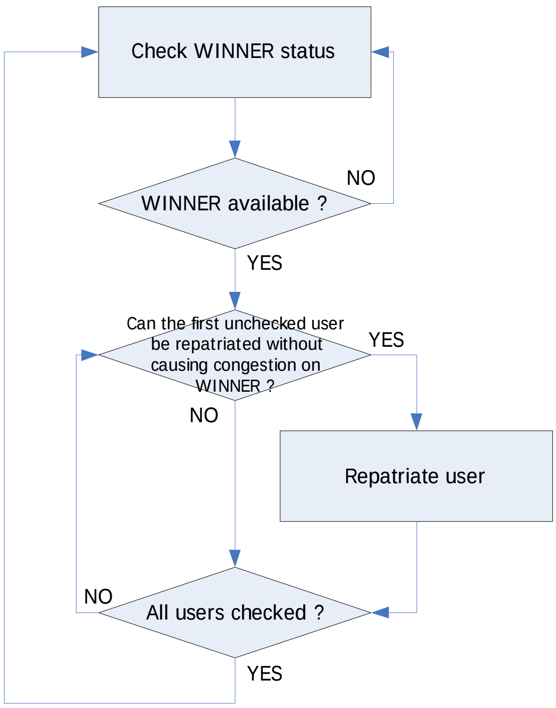

Figure 2-21: Repartition of WINNER Users

In this process, the decision is taken based upon the initial requirements of the application as defined in the application profile. If a user is using simultaneously several applications, he is only repatriated with all his applications. A particular threshold has to be set to identify if the repatriation causes congestion, which may be different from the one used in congestion control.

2.3.4.3 Simulation Results

Different choices have been made to define the environment of simulation. The cells are in theory commonly defined as hexagonally shaped. This is though an ideal representation since in reality the shape of the cells is precisely adapted to the network configuration and determined by the antennas coverage. For our simulations, we have decided to use square-shaped cells. The square shape is not a critical parameter for our model and it is the easiest shape to define in computing language.

Four parameters are used to define each cell: the RAN and mode type, the cell coverage, the cell location and the cell capacity. With this information every cell can be uniquely identified. Besides, we decided to use typical cells for each RAN or mode that is to say that within the same RAN or mode, all the cells will have the same coverage area and capacity value.

Our first assumption is that we use a total of 19 cells in our simulations. These 19 cells are composed of 9 WINNER #1 cells, 6 WINNER #2 cells, 1 WINNER #3 cell, 1 GPRS cell, 1 UMTS cell and 1 WLAN cell.

The choices made in terms of coverage are shown in Table 2-16. We have limited the area of study to the highest coverage area. This means that the whole area of study will be entirely covered by the one WINNER #3 cell. The positioning of the cells is done by placing the upper left vertex of the square-shaped cell in the area of study.

RAN / Mode	GPRS	UMTS	WLAN	WINNER #1	WINNER #2	WINNER #3
Square edge length (a.u.)	8	10	6	6	10	30
Coverage (a.u.²)	64	100	36	36	100	900

Table 2-16: Cells dimension and coverage

Regarding the capacity, the same value for all the legacy RANs' cells has been used, while the value for all the WINNER cells is twice greater.

Then, to determine the total capacity C_{total} of the network, we start with the assumption that we need to host 5 000 users at most within one hour. Using the applications' penetration factors, typical duration and simulated data rate described earlier, we can determine the total capacity value, using the equation below.

$$C_{total} = \frac{N_{U\_\max} x \sum_{i=1}^{N_{SC}} f_i x DR_i x TD_i}{FD}$$

Where N_{u_max} is the maximum number of users (5 000 in our case)
N_{SC} is the total number of service classes (18 in our case);
f_i is the penetration factor of the i^{th} service class;
DR_i is the simulated data rate of the i^{th} service class, in Mbps;
TD_i is the typical duration, or expected download time, of the i^{th} application, in s;
FD is the full duration of the time interval, in s (3600 in our case).

By applying this formula we get the value of 2 625 Mbps for the total capacity of the entire network. Capacity values for each cell depending on the RAN or mode it belongs to are summed up in Table 2-17.

RAN / mode	Cell capacity value (Mb/s)	Number of cells
WINNER #1	150	9
WINNER #2	150	6
WINNER #3	150	1
GPRS	75	1
UMTS	75	1
WLAN	75	1
Entire Network	2625	19

Table 2-17: Cells Capacity Values

The load is calculated independently for each cell. It is expressed as the occupation of the total capacity of the cell, and the load value is then comprised between 0 and 100.

As stated previously, the calculation of load for WINNER has been performed taking only into account the bandwidth metric, in the data rate sense. For simplicity matters the same process has been applied to other RANs. This allows the algorithm to work with a generic formula. Moreover it provides load values that are coherent for the simulation, i.e. the load values for each RAN are comparable since they do not come from totally different calculations.

In order to apply this method, the data rate utilised by each user, depending on the application required, is determined. By adding the data rates of all the users, the load value is obtained, as depicted by Equation 4.2.

$$L_n = \frac{\sum_{i=1}^{N_{ucell\_n}} DR_i}{C_{cell\_n}} \tag{4.2}$$

Where: L_n is the load of the n[th] cell;
C_{cell_n} is the total capacity of the n[th] cell;
N_{ucell_n} is the total number of users running applications in the n[th] cell;
DR_i is the data rate of the i[th] user.

For the simulation a typical delay has been defined for each RAN and for each WINNER mode. The typical delay is the average packet delay happening when the network is in the normal operation phase. Typical delay values are summed up in Table 2-18.

RAN	Typical delay value (ms)
GPRS	200
UMTS	100
WLAN	20
WINNER #1	20
WINNER #2	100
WINNER #3	200

Table 2-18: RANs' typical delay values

This parameter is determined regarding what are the delay requirements of applications that can operate on it and withdrawing an average tendency. It bears no QoS properties but only the big picture about how delay can be included in the algorithms

2.3.4.4 Traffic Load Scenarios

For the simulations we used different types of traffic repartition on the network. Each scenario emulates the traffic for one whole hour. The three scenarios that we chose are Busy Hour (BH), which is a daily period of high network occupation; Normal Hour (NH), which is a period of average network occupation; and Sport Event (SE), which is an unusual period of very high network occupation. By applying these scenarios to our network we will see how the latter performs in different load situations.

The users' repartitions for the different traffic load scenarios are shown in Table 2-19, Table 2-20 and Table 2-21.

Service Class	Number of users per service class	Number of users per group	Penetration factor
1, 2, 5, 7, 9, 11	336	56	6,71
3, 4, 6	18	3	0,36
8	846	141	16,91
10, 14, 15, 16, 17, 18	60	10	1,20
12	1530	255	30,58
13	198	33	3,96
Total	5004	Total	100,00

Table 2-19: Users Repartition for Busy Hour

Service Class	Number of users per service class	Number of users per group	Penetration factor
1, 2, 5, 7, 9, 11	222	37	6,65
3, 4, 6	12	2	0,36
8	570	95	17,09
10, 14, 15	36	6	1,08
12	1032	172	30,94
13	132	22	3,96
16, 17, 18	42	7	1,26
Total	3336	Total	100,00

Table 2-20: Users Repartition for Normal Hour

Service Class	Number of users per service class	Number of users per group	Penetration factor
1, 2, 5, 7, 9, 11	468	78	6,68
3, 4, 6	24	4	0,34
8	1188	198	16,97
10, 14, 15, 16, 17, 18	84	14	1,20
12	2160	360	30,85
13	270	45	3,86
Total	7002	Total	100,00

Table 2-21: Users Repartition for Sport Event

The simulations have been ran for four different values of the congestion thresholds: 70% of the load for the upper value – 65% of the load for the lower value; 75% - 70%; 80% - 75% and 85% - 80%. Moreover, for each TLS, the simulations were performed four times with each threshold value so as to limit the effect of the grouping of users. All these parameters are summed up in Table 2-22.

Busy Hour	Congestion thresholds (lower and upper)	65-70 %	70-75 %	75-80 %	80-85 %
	Number of simulations per threshold couple	4	4	4	4
	Total number of simulations	16			
	Simulation duration	1200 s			
	Number of users	5004			
	Number of groups	108			
Normal Hour	Congestion thresholds (lower and upper)	65-70 %	70-75 %	75-80 %	80-85 %
	Number of simulations per threshold couple	4	4	4	4
	Total number of simulations	16			
	Simulation duration	600 s			
	Number of users	3336			
	Number of groups	108			
Sport Event	Congestion thresholds (lower and upper)	65-70 %	70-75 %	75-80 %	80-85 %
	Number of simulations per threshold couple	4	4	4	4
	Total number of simulations	16			
	Simulation duration	3600 s			
	Number of users	7002			
	Number of groups	108			

Table 2-22: Summary of simulations parameters

We stated before that each group will request an application with respect to the group number. This selection is shown in Table 2-23.

Service Class/ Application requested	Group IDs
1	1, 19, 37, 55, 73, 91
2	2, 20, 38, 56, 74, 92
3	3, 21, 39, 57, 75, 93
4	4, 22, 40, 58, 76, 94
5	5, 23, 41, 59, 77, 95
6	6, 24, 42, 60, 78, 96
7	7, 25, 43, 61, 79, 97
8	8, 26, 44, 62, 80, 98
9	9, 27, 45, 63, 81, 99
10	10, 28, 46, 64, 82, 100
11	11, 29, 47, 65, 83, 101
12	12, 30, 48, 66, 84, 102
13	13, 31, 49, 67, 85, 103
14	14, 32, 50, 68, 86, 104
15	15, 33, 51, 69, 87, 105
16	16, 34, 52, 70, 88, 106
17	17, 35, 53, 71, 89, 107
18	18, 36, 54, 72, 90, 108

Table 2-23: Application request Vs. Group ID

2.3.4.4.1 Results

Here, we present the load and the mean user throughput vs. the congestion thresholds. The values represent the mean of the previous calculated values over the cells of the same RAN/mode. For Busy Hour without QoS algorithms, one set of simulations was made, but since there are no QoS algorithms, no congestion threshold value is defined. Therefore, the presented value is apart from the others.

Figure 2-22 shows that for the WINNER #1 mode, the load increases for all the TLS when the congestion threshold (CT) increases. Nevertheless, except for BH without QoS, the load never reaches exactly the threshold value. Indeed, all the cells might not be fully utilised due to the random positioning of users, so that in average the threshold is not reached. For the BH without QoS, the load represents 90% of the capacity, which is higher than the highest upper CT and would be difficult to handle for the network.

The differences between the TLS values for a given CT are small, which indicates that even in a low-loaded situation, the WINNER #1 cells are much utilised. Hence, when the traffic increases, they cannot be much more solicited. This is due to the fact that whatever the TLS is, the WINNER #1 cells host users with high-rate applications that require many resources. This tends to demonstrate that in the configuration we chose, the WINNER #1 cells should be densely deployed in order to have a very tight coverage. Moreover, the cells should be able to handle more traffic in terms of capacity than the other modes cells.

We can also notice that for the congestion thresholds 75-80 and 80-85, the load for respectively BH and normal hour (NH)_ is higher than the load for respectively sports event (SE) and BH. This can come from the cooperation mechanisms use: the WINNER #1 mode is supported by the other modes, so the load decreases.

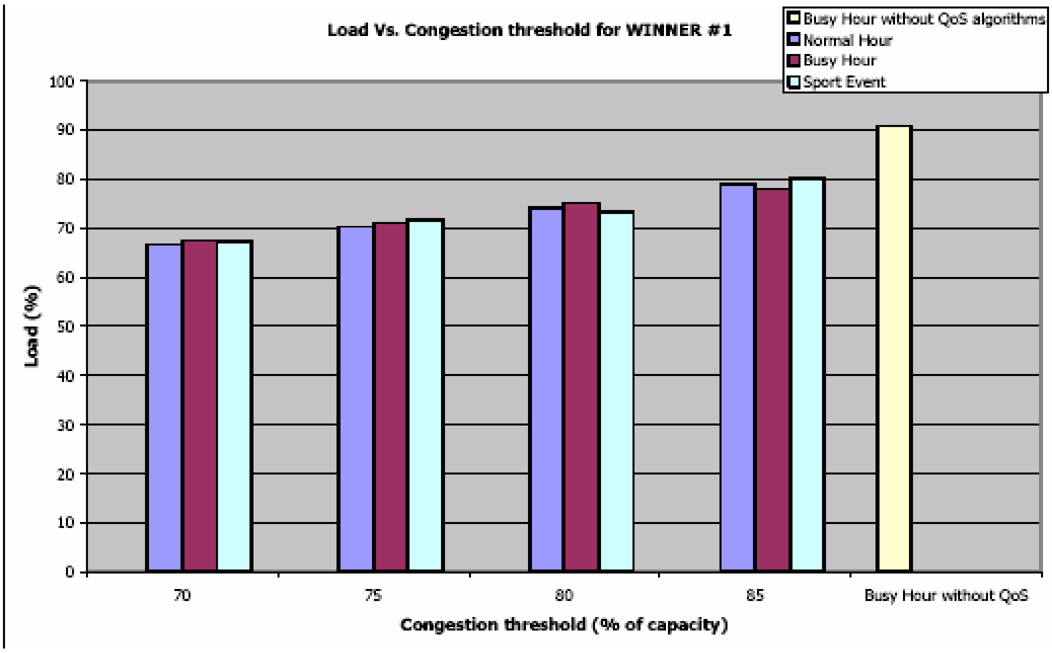

Figure 2-22 Load vs congestion threshold for the applications related to WINNER #1

In Figure 2-23 we observe that the load increases when the CT and the traffic increase. We can notice that the higher the threshold, the higher the relative difference between NH and the other TLSs. It indicates that for NH, the WINNER #2 cells do not experience congestion and are under-utilised, whereas for the other TLSs, the CTs are almost reached. Indeed, the WINNER #2 cells do not accommodate very high-rate applications, so they are less solicited when the traffic is low. But as soon as the traffic increases, more use is made of them for supporting the WINNER #1 mode with the high- or medium-rate applications that WINNER #1 no longer accepts. For SE, the load values are very close to the CT except for the CT 85. Indeed, the CT being higher, the cells can accommodate more users and the congestion control mechanisms are less solicited in this case. For BH without QoS, the load value reached is once again very high (about 89%) which indicates the efficiency of the QoS algorithms.

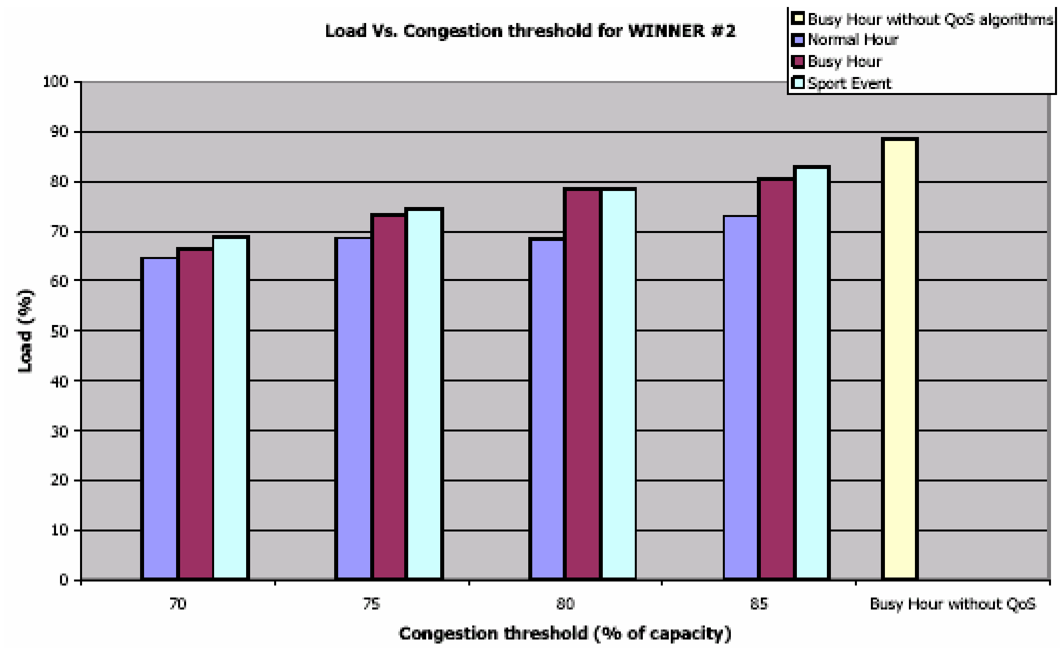

Figure 2-23 Load vs. congestion threshold for the mode WINNER #2

Figure 2-24 shows that the load for NH is very low (under 50%), which indicates that the cell is under-utilised due to the fact that most of the traffic is handled by WINNER #1 and #2. It also comes from the fact that users placed in this cell mostly use low-rate applications for this TLS. However, the load increases for BH and reaches the CT for SE, which tends to prove that this cell is used as a support for the others. The WINNER #1 and #2 cells accommodate users with very-high- and high-rate applications, and the other users are placed in WINNER #3. For BH without QoS, the load is still high but not as much as previously (about 82 %) and is even exceeded by the SE load for the CT 85. This confirms the idea of the role of support played by WINNER #3.

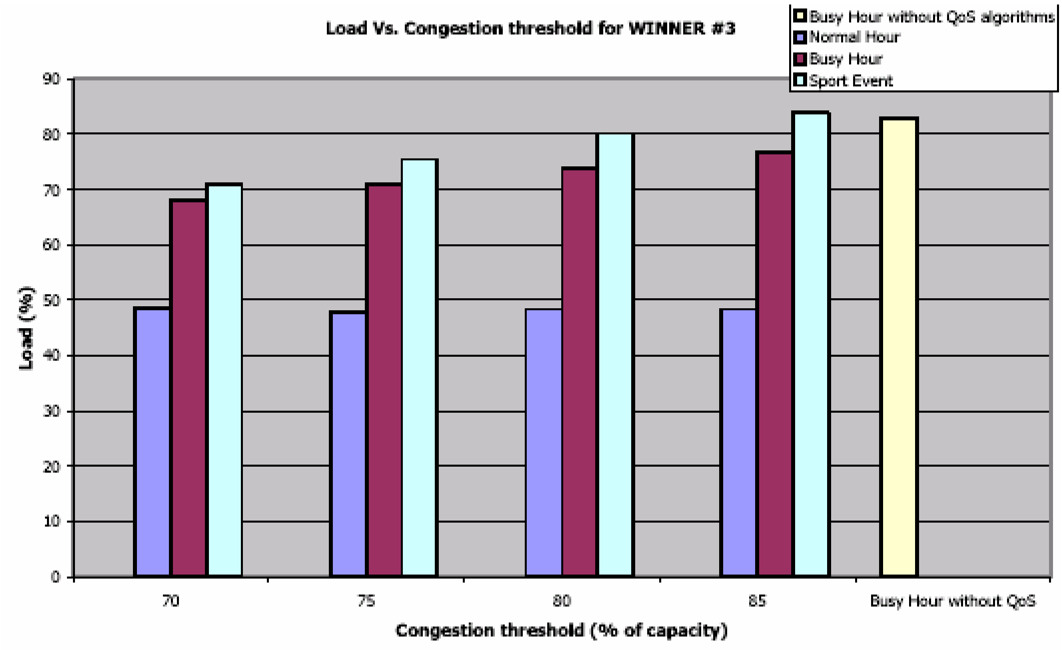

Figure 2-24 Load vs. congestion threshold for the mode WINNER #3

Figure 2-25 shows that the legacy RANs are very few utilised in NH since the load is inferior to 30%, which indicates that the WINNER cells are used in priority. For BH and SE, the tendencies are the same as for WINNER #3: when the traffic load increases, the legacy RANs are much more utilised, although the load values do not reach the CTs. This tends to show that the legacy RANs are also used to support other networks. However, the load barely changes for BH and SE when the CT changes, which shows that the traffic is handled. Of course, the legacy RANs are not compatible with many applications, which can also explain their limited utilisation especially in BH.

We can conlcude that in the case of low traffic situation, most of the traffic is handled by WINNER #1 and #2. When changing the TLS, the other networks are used for support purposes. This indicates that the cooperation mechanisms are completely needed since their activation ensures that the traffic repartition is much more homogeneous even if the legacy RANs cannot support many applications.

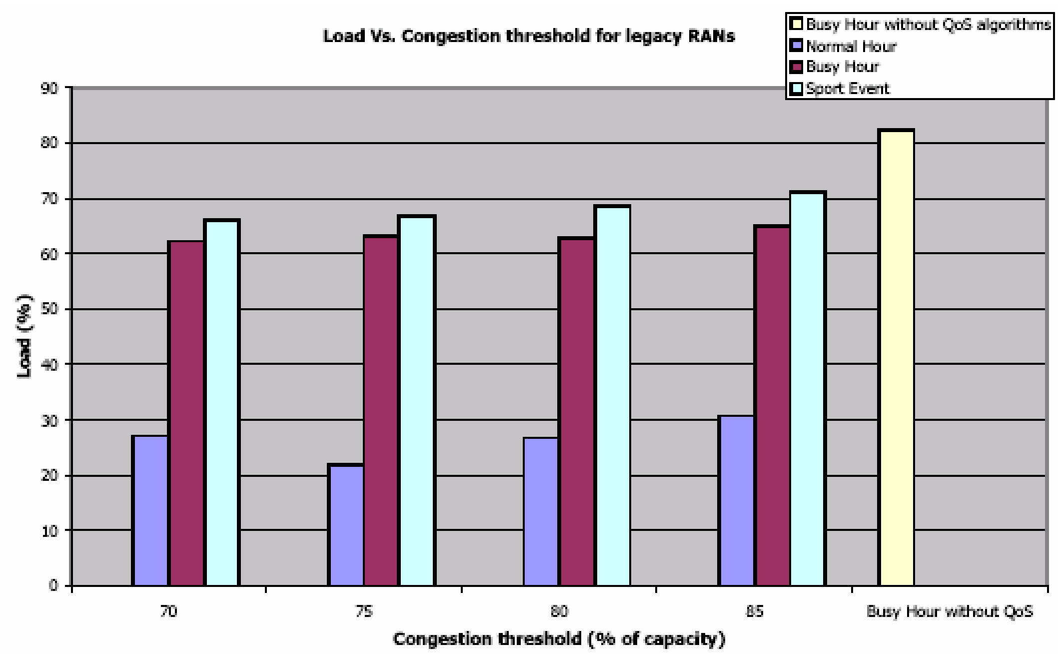

Figure 2-25 Load vs. congestion threshold for the legacy RANs

Figure 2-26 to Figure 2-29 show the results for the mean user throughput for each traffic load scenario. We can see from Figure 2-25 that for NH, the Mean User Throughput (MUT) remains almost the same, which is consistent with the results for the load. For BH, the MUT tends to decrease between CT 70 and CT 85. It indicates that the QoS renegotiation processes are more applied when the CT is higher than 70.The load of the cells WINNER #1 increases, while the percentage of dropped users decreases. So the MUT is diminished but there are less users dropped, which tends to show that the QoS mechanisms are better employed when the CT is higher than 70. This tendency is confirmed by the SE results, where the MUT value is high for CT 70 and decreases for the other CTs.

For BH without QoS, the MUT is relatively high, but for UP1 and UP2, the percentage of connected users is very low, and the percentage of rejected users very high. So the remaining connected users use more resources since they are less numerous.

The MUT values are globally high (between 14 000 and 34 000 kbps) which confirms that the users connected in the cells WINNER #1 use very high- and high-rate applications. In Figure 2-27, we observe that for NH, the MUT is stable for the three first CTs but decreases at CT 85, which is consistent with the results for the load. Once again, the QoS decrease processes are more utilised since the load increases. For BH and SE, the MUT experiences globally a decrease when the CTs increase, once again in parallel with the load increase. The MUT for BH without QoS is lower than for the cells WINNER #2, but the percentages of connected and rejected UP1 and UP2 users have the same behaviour than for WINNER #1. The MUT values are lower than for WINNER #1 (between 7 000 and 22 000 kbps), which is consistent with the defined UPs: WINNER #2 hosts medium- and high-rate applications.

Figure 2-28 shows that for NH the MUT is stable, which corresponds to the load results of Figure 2-24. With the increase of traffic, the MUT increases too, which confirms that in low-traffic situation, WINNER #3 hosts users using low rate applications whereas in BH and SE, it supports the other networks by hosting also users with medium-rate applications. For SE, the MUT decreases regularly. This is the same tendency as before, i.e. the algorithms are more utilised with high CT.

For BH without QoS, the MUT reaches values comparable with the values obtained when the algorithms are used. Indeed, the percentage of UP3 connected users is almost the same with and without QoS algorithms, which means that these users have the resources they need considering the capacity of the cells they are placed in.

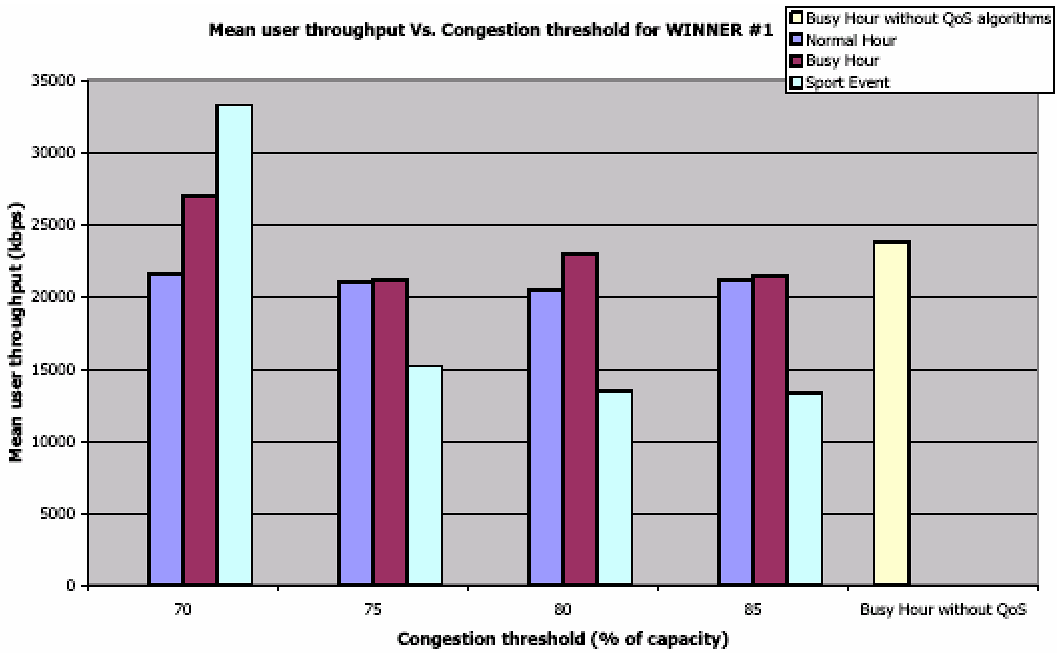

Figure 2-26 Mean user throughput vs. congestion threshold for the mode WINNER #1

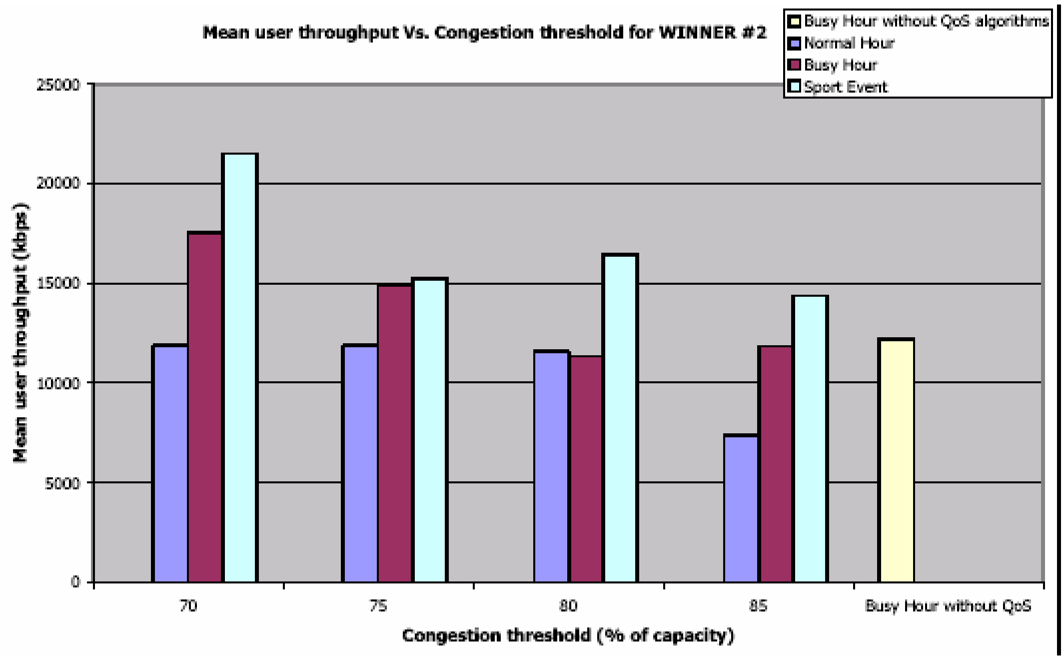

Figure 2-27 Mean user throughput vs. congestion threshold for the mode WINNER #2

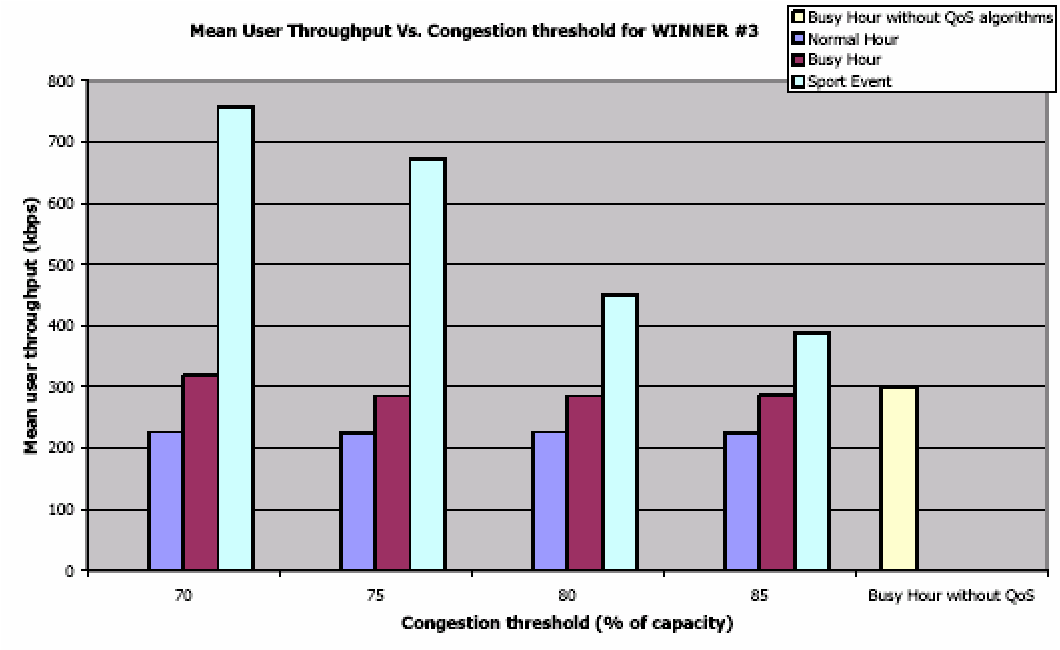

Figure 2-28 Mean user throughput vs. congestion threshold for the mode WINNER #3

In Figure 2-29 we can see that for the three TLSs, the MUT increases except for SE at CT 85 where it decreases, probably because the QoS renegotiations are more applied. It tends to show that for the legacy RANs, the tendency is the opposite of what happened in the WINNER modes. It confirms the action of mechanisms of cooperation especially at CT 85: if for SE the increase was proportional to the increase observed for NH and BH, the cells could not handle the load generated, so we can suppose that the MUT value embeds the QoS decrease effect. For BH without QoS, the value is high (14 000 kbps whereas it is between 1 000 and 10 000 in the other cases). It corresponds probably to the use of the WLAN cell. The users placed in this cell are UP1 users, so as stated before, they can have many resources since they are very few connected users and many rejected.

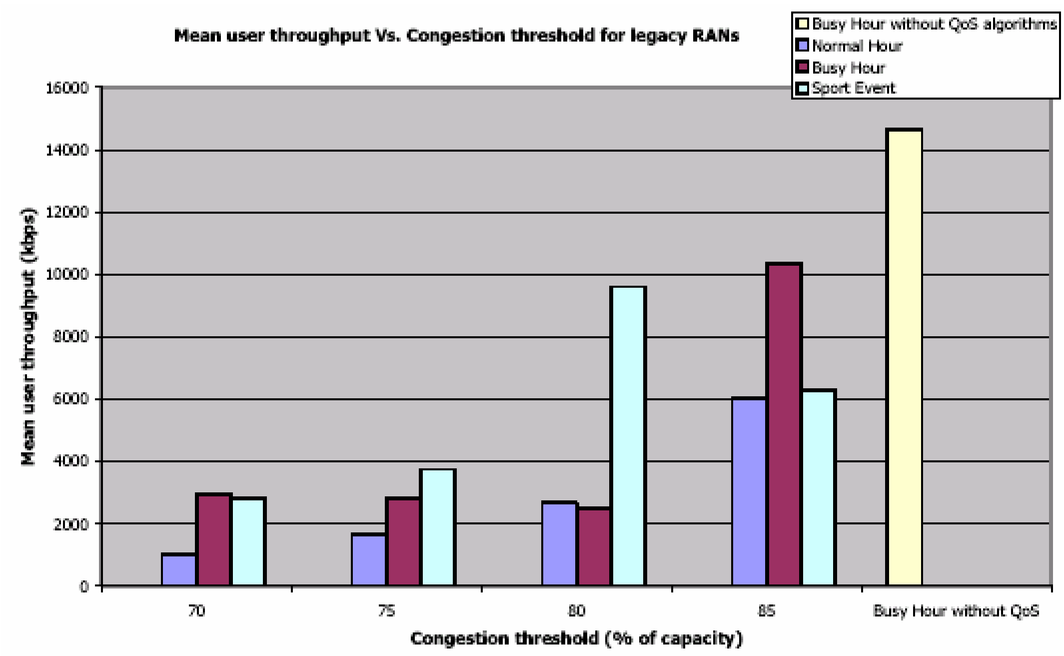

Figure 2-29 Mean user throughput vs. congestion threshold for the legacy RANs

It must be mentioned that the presented results are based on a simplified approach of user and application prioritisation and load computation. Although, the results cannot be used for a finalised conclusion for the number of blocked, dropped and connected users, they are indicative for the performance of the cooperation mechanisms and their role for the support of intermode and inter-RAN cooperation. Further work will include and improved methododlogy for user and application prioritisation as well as a more precise (to include the performance specifics of the WINNER RAN) approach for calculating the load of the network.

3. Cooperation in the user plane

In order to achieve high reliability for multi-hop communication in combination with mobility in the wide area and metropolitan area deployment scenarios (RRM modes), more advanced ARQ protocols are required at the link layer than the ones used in networks with only single hop communication. Except from data loss due to bit errors over radio links, data may also be lost in RN on the way between a sender and a receiver. The individual links in a multi-hop transmission will most likely have different characteristics. The datarate between a RN and a base station (BS) may be higher than the datarate between the RN and a user terminal (UT). The radio quality may be uneven over the links. Data loss may be caused by handover or RN switch between BSs. Error recovery only on a hop-by-hop basis is not sufficient to recover from all types of errors that may occur. In order to provide reliability to higher protocol layers, such as TCP, ARQ is also required over all wireless hops involved in a transmission, i.e., between the nodes at the edges of the wireless network, the UT and a node connected to the fixed part of the network. The node in the fixed part is typically the RANG. In some situations, as described in [WIND32], a BS could serve as termination point. In the following, the RANG is assumed as termination point.
In this section, three ARQ approaches intended for multi-hop communication in the WINNER system are presented. In section 3.2, the mapping of protocol entities to logical nodes are discussed. The ARQ approaches are described in [WIND32] and [WIND35] from WINNER phase I. The first approach, the layered ARQ approach, is the straight forward solution that uses two ARQ layers, one hop-by-hop and one edge-to-edge. The layered approach is, for example, specified for GPRS to combat handover losses [ETSI98]. The second approach, the Relay ARQ approach [Wie05], is based on a new link layer concept. The idea is to have one single protocol that provides both hop-by-hop and edge-to-edge error recovery. The third and last approach, the Multi-hop ARQ [Lott05], is a variant of the layered ARQ approach. The main idea here is that it should be possible to use legacy link layer protocols to communicate with UTs.

3.1 Proposed ARQ approaches

3.1.1 Layered ARQ

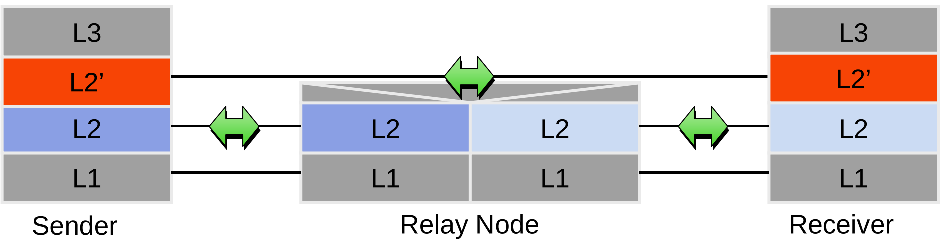

Figure 3-1: Layered ARQ approach

In the layered ARQ approach, illustrated in Figure 3-1, two levels of link layer ARQ protocols are used. One ARQ protocol operates over each hop (L2), e.g., a hybrid ARQ protocol. In some systems, e.g., HSDPA and the 3G Long Term Evolution (3G LTE), this layer is specified as the MAC sublayer in L2. Another ARQ protocol (L2') operates between the RANG and the UT at the edges of the wireless network. This layer is normally operated at the Radio Link Control (RLC) sublayer of L2. The layered ARQ approach works for two or more wireless hops. The hop-by-hop protocol is responsible for recovering from transmission errors that occur over a link between to adjacent nodes. The edge-to-edge protocol is responsible for retransmissions if data is lost due to handover between RNs.

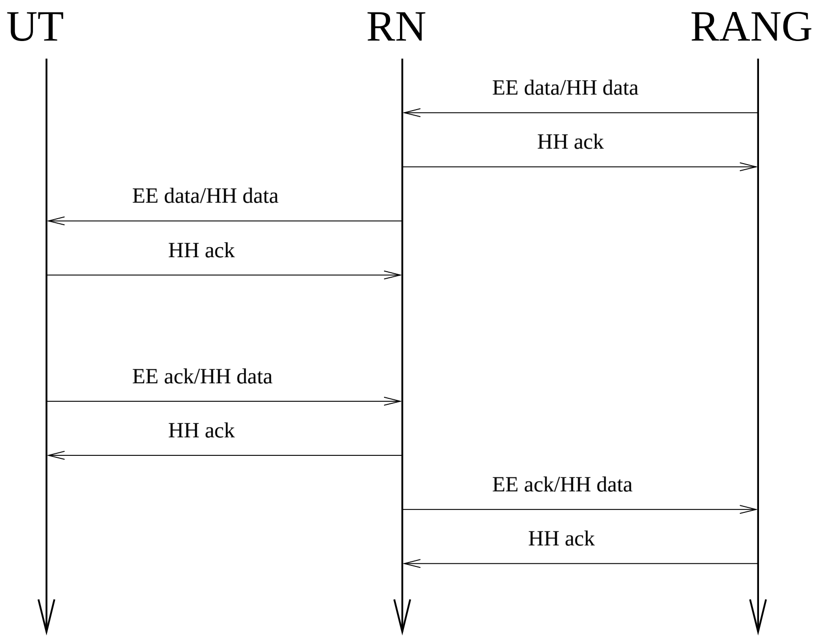

Figure 3-2: Layered ARQ approach: successful transmission

The layered ARQ approach is flexible. Different hop-by-hop protocols can be used over the hops. The frame size of the hop-by-hop protocol can be optimized for the conditions of the individual radio links. One edge-to-edge frame may be contained in one or more hop-by-hop frames. The sequence diagram in Figure 3-2 shows a successful transmission in downlink of an edge-to-edge frame (until the edge-to-edge protocol transmits an acknowledgment) from the RANG to the UT. In this case there is a one to one mapping between edge-to-edge and hop-by-hop frames.

Having ARQ protocols on two levels may, however, result in negative protocol interaction. It is hard to configure protocol parameters, such as the retransmission time out values to avoid concurrent retransmissions and waste of radio resources due to link under-utilization. If the retransmission timer of the edge-to-edge protocol expires too early, the protocols may unnecessarily retransmit the same data. A higher time out value may, on the other hand result in link under-utilization. If data is lost due to handover between RNs, the links may be under-utilized until the retransmission timer of the edge-to-edge protocol expires. The second link may be under-utilized if the hop-by-hop protocol provides in-order delivery and frames arrive out of order to the RN, since the RN reassembles out of order frames in sequence number order before they are passed to the protocol entity of its second link.

3.1.2 Relay ARQ

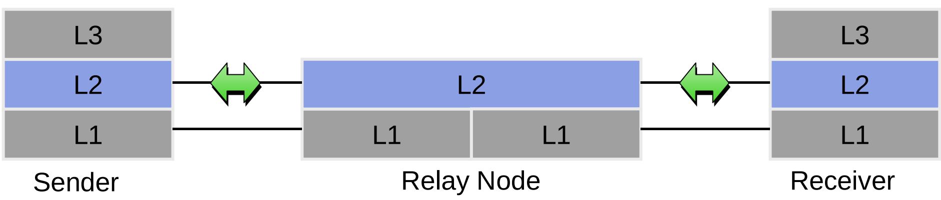

Figure 3-3: Relay ARQ approach

The Relay ARQ approach presented in [Wie05], combines the functionality of the hop-by-hop protocol and the edge-to-edge protocol into one protocol, as illustrated in Figure 3-3. The Relay ARQ protocol entity in a RN integrates the operations over both the links that are interconnected by the RN. The same frames, and, hence, the same sequence numbers, are used over all links between the RANG and the UT. Three types of status messages are used, positive acknowledgment (ACK), negative acknowledgement (NACK) and relay acknowledgement (RACK).

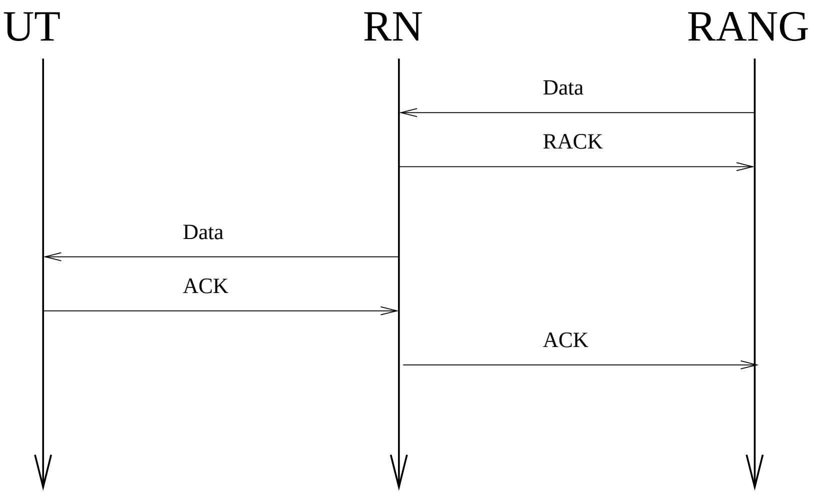

Figure 3-4: Relay ARQ approach: successful transmission

Relay ARQ is intended for communication over two or more wireless hops. In the following, we use the two hop scenario in Figure 3-3 as an example to describe Relay ARQ. Furthermore, downlink operation is assumed. In Figure 3-4 a successful transmission is illustrated. The RN keeps state information about frames from the RANG and about acknowledgements from the UT. The RN transmits a RACK in response when it receives a frame (from the RANG) which is not yet acknowledged by the UT. The RACK corresponds to a hop-by-hop ACK in the layered ARQ approach. At the RANG, a RACK indicates that the RN has taken over the responsibility for the frame. The RANG keeps the frame in the transmission buffer until it receives an ACK from the UT. The ACK from the UT corresponds to an edge-to-edge acknowledgement in the layered approach. If the RN fails to deliver the frame to the UT, e.g. due to handover between RNs, then the node before, in this case the RANG, takes over the responsibility for retransmitting the frame. The original sender has the ultimate responsibility for the frame.

3.1.3 Multi-hop ARQ

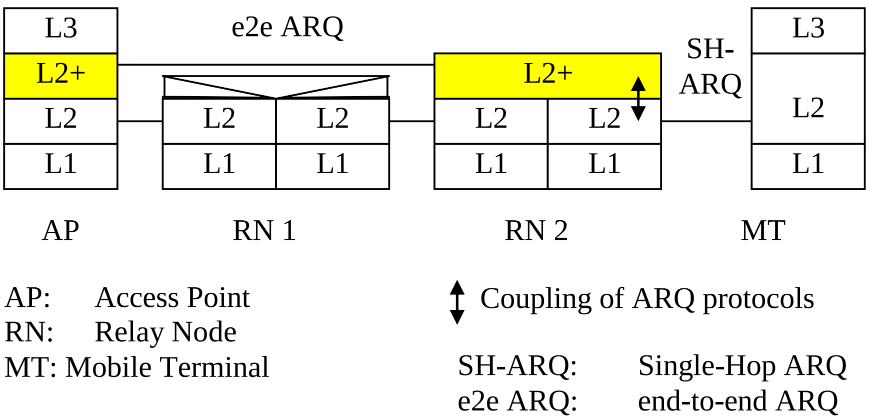

Figure 3-5: Multi-hop ARQ approach

In [Lott05], an ARQ approach for more than two wireless hops is presented. As in the layered approach, two levels of link layer protocols are used. Instead of terminating the edge-to-edge protocol in the UT, the edge-to-edge (L2’) protocol operates between the RANG and the RN closest to the UT, as illustrated in Figure 3-5. Any hop-by-hop link layer protocol may be used between the last RN and the UT. This enables inter-operability with UTs running legacy protocols. The edge-to-edge protocol in the last RN is tightly coupled with the hop-by-hop protocol that is used over the link to the UT.

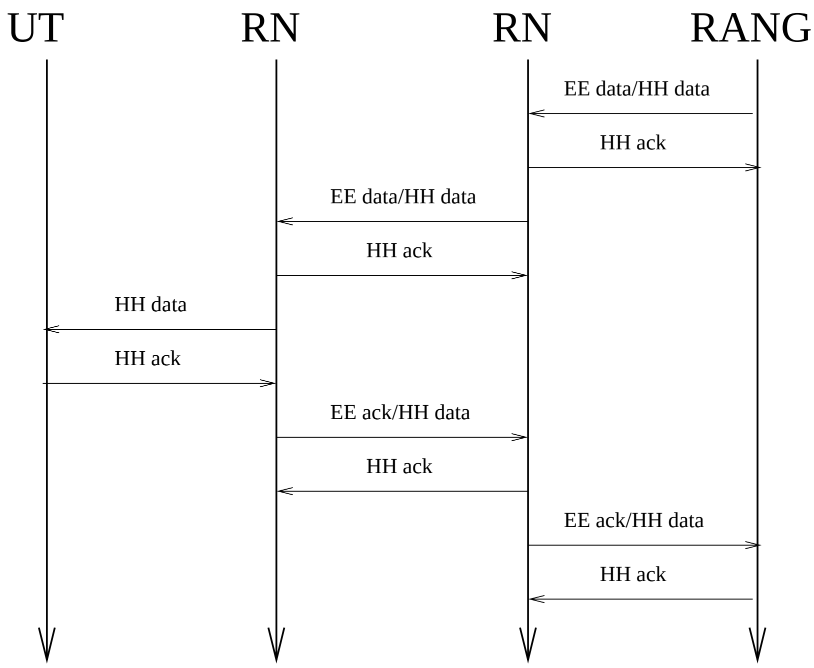

Figure 3-6: Multi-hop ARQ approach: successful transmission

In downlink operation, the edge-to-edge protocol acknowledges data first after it has reached its final destination, as shown in Figure 3-6, and in uplink operation the last RN does not transmit a hop-by-hop acknowledgment to the UT until it has received an acknowledgement from the RANG.

The main advantage with this approach is that it allows UTs to run legacy link layer protocols. On the other hand, the edge-to-edge protocol in the last RN has to be tightly coupled with the hop-by-hop protocol used over the link to the UT. This may require one variant of the edge-to-edge protocol for each hop-by-hop protocol (new and legacy protocols) that should be supported. Computational complexity and memory requirement in the UT is probably lower than with the other approaches. This approach has the same disadvantage as the layered approach, that of potential negative protocol interaction.

3.2 Protocol entities and logical nodes

The WINNER system may have L2 ARQ on three functional layers, RLC-g, MAC-g and MAC-r [WIND32]. Edge-to-edge ARQ is provided by RLC-g, which consists of mode independent functions. Hop-by-hop ARQ is provided by MAC-g and MAC-r. The MAC-g protocol is mode independent. According to [WIND32], MAC-g may consist of Hybrid ARQ, or at least parts of Hybrid ARQ. The MAC-r protocol provides mode dependent functions.

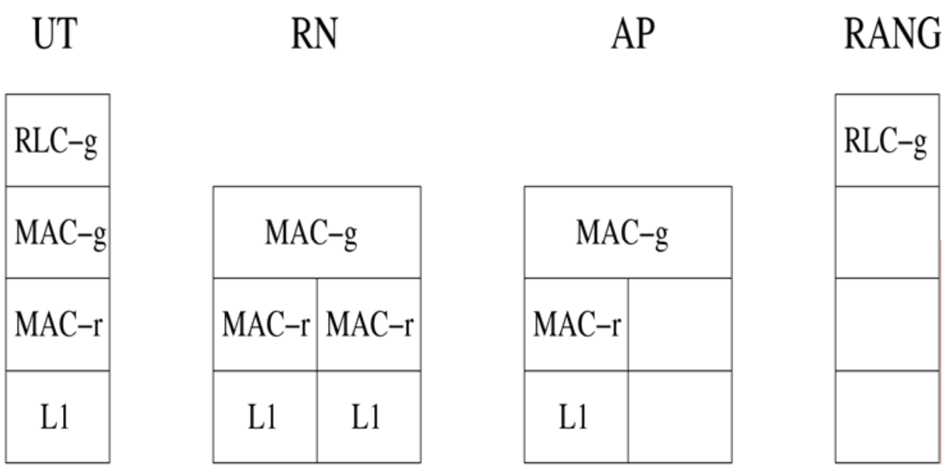

Figure 3-7: Layered ARQ approach: protocol entities and logical nodes

The mapping of RLC-g, MAC-g and MAC-r to logical nodes for the layered ARQ approach is illustrated in Figure 3-7. The hop-by-hop link layer protocols, MAC-g and MAC-r, are terminated in the UT, RN and BS. The edge-to-edge protocol, RLC-g, is terminated in the UT and in the RANG, in most cases. In [WIND32], (sec 2.4.2.4) it is suggested that the termination point could be moved from the RANG to the BS, if there is only handover between RNs and no handover between BSs, since efficient edge-to-edge error recovery only requires that the edge-to-edge protocol terminates one hop away from the node that is handed over.

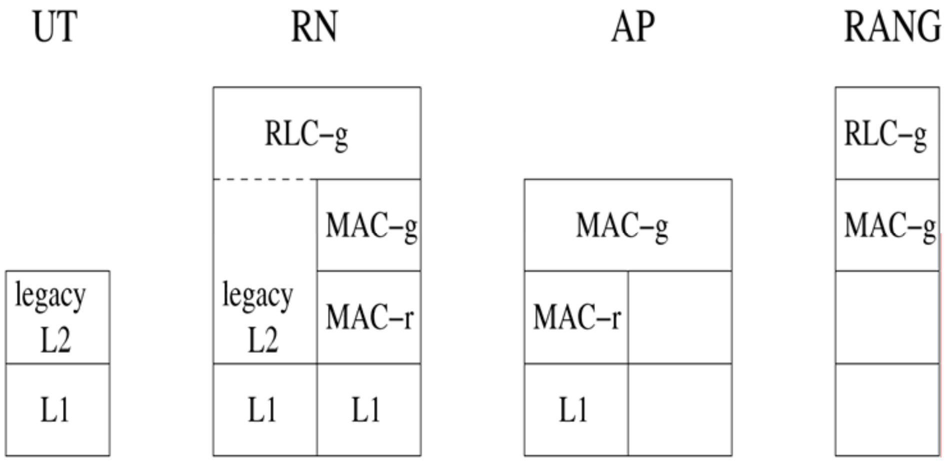

Figure 3-8: Multi-hop ARQ approach: protocol entities and logical nodes

For the multi-hop ARQ approach [Lott05], RLC-g, MAC-g and MAC-r could be mapped onto logical nodes as indicated in Figure 3-8. The multi-hop ARQ approach requires that the edge-to-edge protocol, RLC-g, terminates in the last RN and in the RANG. The hop-by-hop protocols, MAC-g and MAC-r, operate between RNs and between BSs and RNs. The last RN should run the same legacy hop-by-hop protocols that are used in legacy UTs. The legacy hop-by-hop protocols in the last RN should be tightly coupled with the edge-to-edge protocol. This implies that RNs should have two modes of operation, one if it is the last RN and one if it is communicating with RNs and BSs.

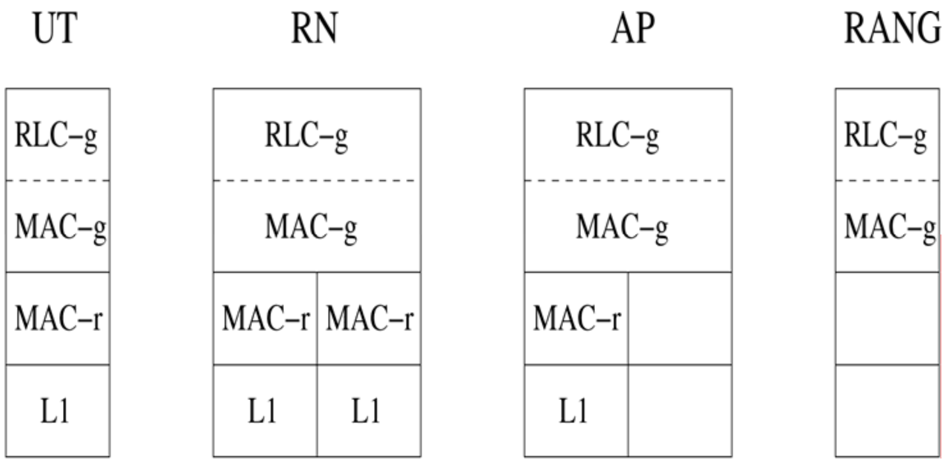

Figure 3-9: Relay ARQ approach: protocol entities and logical nodes

Figure 3-9 shows how the Relay ARQ approach [Wie05] could be mapped onto protocols and logical nodes. Hop-by-hop protocol functionality is required in the UT, RN and BS, but the Relay ARQ protocol should terminate in the RANG. The Relay ARQ protocol entity could consist of the mode independent protocols, RLC-g and MAC-g. The combination of edge-to-edge and hop-by-hop ARQ functionality, would then be provided, since RLC-g represents edge-to-edge functionality and the MAC-g protocol may provide hop-by-hop functionality with Hybrid ARQ.

3.3 Conclusions

Although from previous work that a stable lower layer ARQ mechanism releases the burden of higher layer ARQ, it is still not obvious which ARQ approach that will best meet the requirements of the WINNER system according to the on-going research in the lower layers. The ARQ approaches therefore have not yet been evaluated in detail and compared with each other. In [Wie05], which describes the Relay ARQ approach, some initial simulations and a comparison with the layered approach are presented. Concurrent retransmissions and link under-utilization are probably easier to avoid with the Relay ARQ approach. Inter-operability with legacy link layer protocols is only provided by Multi-hop ARQ.

4. Measurements and triggers for cooperation mechanisms

4.1 Measurements

Measurements are essential inputs for RRM algorithms and therefore mechanisms to configure, perform and report measurements must be defined for the WINNER system. Measurements affect to the design of the WINNER PHY, MAC and RLC (or RRC) from the 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 ACS/SRRM (or/and cooperative RRM) entity a set of measurements for handover and other RRM functionalities:

- **Received signal strength**, **Interference level and C/I 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 either terminals (UT), base stations (BS) and the relay nodes (RN), 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 RANG 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, RANG.
- **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 D4.3 section 3.3.3, for WINNER RAN the input for cell load are presented in this document section 6.2, and should be performed by ACS and RANG
- Terminal velocity and terminal location. As minimum requirement the system should know to which BS the UT will be attached an to know the coverage area of the serving BS, a more detailed position determination should b**e** performed by the ACS**,** using received signal strength measurements or satellite measurements (GPS)

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

- BS/RN measurements:
  - Received Total 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
  - Cell load (at BS level).- Percentage of occupied chunks over total

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

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

The selection of the most appropriate WINNER cell, mode or even wireless system also should be based on the in the previous parameters that the current mode or RAT and possible target mode or RATs could offer to the terminal, and also it will be taken into account other information elements:

- **Network capabilities**. The legacy RANs can not support adequately many of the services that can be offered by the B3G networks, but also B3G networks with limited channelization can provide a limited subset of services. The information on network capabilities should be available (datarate, delay, etc)
- **Terminal capabilities**. The B3G terminals will be classified by its performance in different classes. It is expected that some terminals will have limited performance to achieve some goals as reduce size, battery duration, lower cost, etc.
- **User preferences**. The user could select some characteristics, e.g pre-selecting the network that will offer lower cost per transmitted bit.
- **Operator classification of users**. The operator could offer different classes of subscription, bronze, silver and premium users, with different level of performance
- **Other high level parameters**. The architecture should be enough flexible to accommodate other information elements

Two approaches are possible to design the WINNER multi-mode terminal, able to measure other RATs:

- The dual receiver approach. A dual receiver can use separately dedicated receivers for each source at the same time. However, there are duplication of hardware components might increase the cost and size of the mobile equipment. In addition it is not optimal as each receiver is not busy during all the communication periods.
- The scheduling approach. In the case of the scheduling approach, a single receiver is capable of alternating operation on the resources. This approach has the following advantages: lower cost and size in the hardware at the receiver level. This is the primary option.

In fact, both approaches can deploy the ARMH (Adaptive radio multi-homing) approach to provide system capacity gain [Luo03][Luo04]. Due to the high flexibility and scalability the resource scheduler provides, the first approach gives a relative performance. However, system realization is a trade-off between the cost and the performance.

The WINNER UT, BS and RN should perform measurements, and also it should be establish the mechanisms to report these measurements to the UT/RN/ACS. These measurements should be triggered periodically and on demand.

The WINNER 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 entity in WINNER, associated to the ACS, should have the possibility to know to what possible cells of legacy system could handover to have a fast intersystem handover, when necessary.

4.2 Triggers

The network/mobile has the responsibility selecting the “best” RF channel to use. However, there are two issues to address. First, the network/mobile must possess the necessary data to make a valid decision. The second issue is that the mobile is required to frequently compare the current channel with possible alternate candidates, i.e. it has to make measurements. In general, a logic separation between information gathering and evaluation, respectively decision taking has to be taken. The usual case will be that some entities mainly located within layer 1 and 2 will collect the respective information, i.e. measurements, and forward them to one central instance, that takes over the evaluation. Incoming L1 measurement, e.g. signal strength, or resulting Block Error Rate (BLER) or number of counted negative acknowledgements (NACKs) will be responded by firing a trigger. Triggers are mechanism that indicate changes of setup or surrounding conditions, and they are essential for HO decisions

The first group consists of triggers that necessitate handover and therefore if a handover does not take place the call will be dropped. The second group contains triggers that can cause a handover but if not performed it won't result in a call being dropped.

A number of triggers are expected to be used for the initiation of an intermode or intersystem handover. However, the importance of those triggers will not be the same. Herein, we attempted to categorize the anticipated triggers and measurements in two main groups. The first group consists of triggers that necessitate handover and therefore if a handover does not take place the call will be dropped. The second group contains triggers that can cause a handover but if not performed it won't result in a call being dropped.

Triggers on current mode / RAN that necessitate handover:
- Signal strength
- Interference level
- BLER
- SINR
- Increase/decrease of UT velocity
- New service request/release
- Cell congestion

Triggers on current mode / RAN that can cause handover: but do not necessitate:
- Terminal location
- Current cell load is higher than target cell load
- User preferences (price, operator) can be fulfilled on target mode but not on current mode (or RAN)
- Higher datarate reachable on target mode than on current mode (or RAN)
- Higher QoS reachable on target mode than on current mode (or RAN)
- User's class of service (bronze users on UMTS/ gold users on WINNER for instance)
- Operator's policy concerning service (voice on GSM for instance) service availability
- QoS violation

Although a handover might be initiated it might not be completed. Below we list some of the foreseen reasons for such a handover rejection.

Reasons for rejecting handover at the BS:
- Target RAT belongs to a not acceptable operator
- New target cell load/target new cell capacity
- Drop call rate too high on target cell
- Blocked call rate too high on target cell
- Handover failure rate too high on target cell
- QoS violation on target cell.

Reasons for rejecting handover at the MT (assuming the BS doesn’t have the information):
- MT capabilities
- QoS violation

4.3 WINNER MAC frame definition

The different WINNER handover functionalities (inter-RAN, intermode and intramode) are strongly related with the PHY layer measurements, as WINNER is based on OFDM, the power measurements rely on pilots, that are used not only for RRM functionalities, also for several functions as connection setup, synchronization, power control and CQI measurements.

In the current design of the WINNER super-frame (SF), is composed of a preamble and a number of frames, currently 8, composed each one by a UL and DL sections. Figure 4-1 presents the current structure of the TDD super-frame, for further details about the super-frame design, the interest reader is referred to [WIND76] [WIND210]. Basic unit of information that could be assigned in the UL or DL to a Radio Access Point (RAP) or UT is called chunk consisting of nsub adjacent subcarriers and nsymb consecutive OFDM symbols.

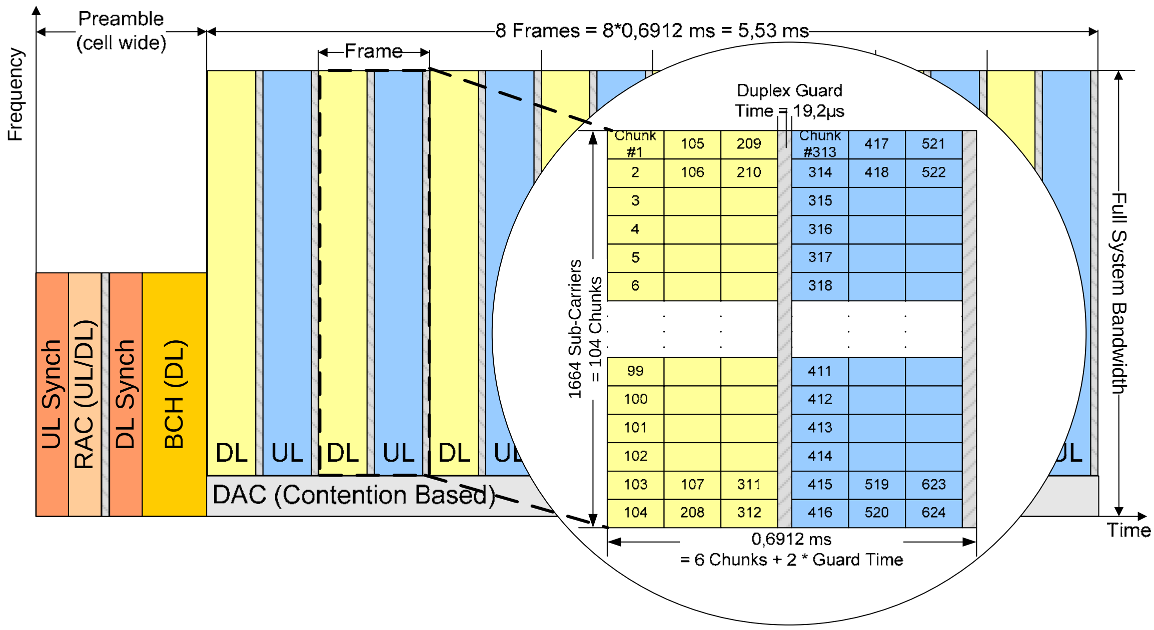

Figure 4-1: WINNER MAC super-frame for TDD

At the beginning of each super-frame there are **two synchronization slots (UL/DL)**, Self-organizing synchronization of terminals and network nodes, as described in [WIND23] can be used on a super-frame basis by this design.

In the **downlink synchronization slot**, each base station/relay node transmits on four OFDM symbols. The first, the *T-pilot*, is used for coarse synchronization. In the remaining three symbols, each BS transmits on two (random) adjacent subcarriers, with the others set to zero. On reception, they are used for updating the UT synchronizations.

In the **uplink synchronization slot**, the **first slot** of three OFDM symbols the next super-frame is used from synchronization of neighbouring BS and relay nodes. Here, all terminals transmit at the same time and frequencies on the two adjacent subcarriers that were received strongest, i.e. those that were used by the BS/RN closest to them (and therefore as first approach could be the best BS/RN). In the FDD mode, the UL synch. slot is in the UL super-frame and the DL synch. slot is in the DL super-frame.

Regarding power measurement only the DL synchronization pilots could be used, but not the UL synchronization pilot because all the UT terminals in cell will transmit at the same time, and therefore the contribution from each UT can not be distinguished. Furthermore, the two randomly select pilots selected for the DL synchronization, will use a frequency in all the band used by the different chunks, not necessarily the comprised by the band of the chunk in which the measurements will be taken. A more precise information on the chunk received power will be provided by the embedded pilots included in the chunks. Two kind of embedded pilot have been defined:

Dedicated pilots.- May be required for user –specific pilots transmit data processing . Two subclasses have been defined: Full band dedicated pilots and Chunk-specific pilots.

- **Common pilots.-** These pilots use the same amplitude and phase ***over all chunks***. Three subclasses have been defined:
- **Common pilot per cell/sector**. This pilot is transmitted omnidirectionally (not subjected to beamforming). It is used for support **mobility-related** functions. It allows obtaining an unweighted channel coefficient H.
- **Common pilots per antenna.** Used to obtain matrix channel H any combination of transmit and receive antennas
- **Common pilots per beam.** Used to obtain the effective channel per beam.
- Figure 4-2 shows the pilot distribution per chunk. The TDD mode uses 2 pilots for Adaptive a Non Adaptive data flows. The FDD mode uses 6 pilot for Non Adaptive mode and 4 pilots for Adaptive mode

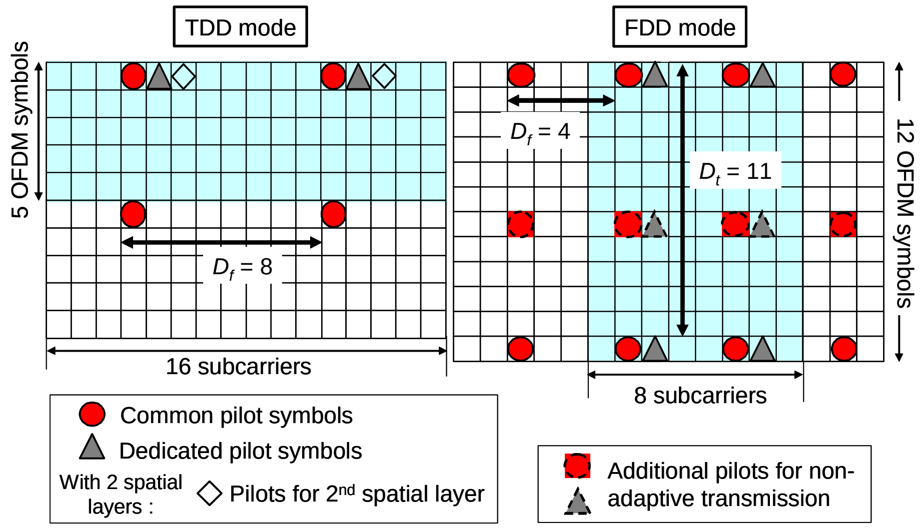

Figure 4-2: Pilot grid in frequency and time domain for TDD and FDD modes

4.4 Inputs for WINNER cell load computation

Besides the measurements for handover and other RRM functionalities specific information should be available for WINNER load computation at MAC, RLC and SLC (service level controller) levels:

MAC user plane (associated to the RANG and/or BS/RN)

- Number of used chunks divided by the total number of chunks available to data transmission.
- Retransmission rate. Number of retransmitted MAC packets divided by the total number of transmitted packets.
- Interference level in chunks
- BLER

SLC

- Buffer occupancy. Total, average and variance of the buffer occupancy (of the different buffers).
- Delay. Total, average and variance of the packet delay.
- Available bandwidth (throughput)

RLC control plane (renamed RRC)

- Queued users. Number of users in the admission control queue.
- Request for short term resource assignment
- Request for proactive routing (more than two hops)

4.5 Measurements for location determination

Positioning in wireless networks became very important in recent years and services and applications based on accurate knowledge of the location of the UT will play a fundamental role in the future wireless systems. In addition to vehicle navigation, fraud detection or automated billing, it is necessary that all wireless service providers have to deliver the location of all emergency callers with specified accuracy. Besides these user specific service on demand applications, also for system side implementations accurate knowledge about the UT position becomes more and more interesting. Especially for LBH, LBVH or RRM the estimates of the UT location can be an input to improve the performance [STK05].

A solution is the determination of the UT location by exploiting the already available resources of a cellular network provided by the WINNER RAN. In cellular network based positioning the localization process is generally based on measurements in terms of Time of Arrival (ToA), Time Difference of Arrival (TDoA), Angle of Arrival (AoA), and/or Received Signal Strength (RSS), processed by the network or UT [GG05].

Another solution is based on Global Navigation Satellite Systems (GNSSs) [PS96]. UT localization using GNSSs such as the available Global Positioning System (GPS) or the future European Galileo system deliver very accurate position information for good environmental conditions, i.e., for direct Line of Sight (LoS) access to several satellites the achievable accuracy can be very high (e.g. WA). Nevertheless, the performance loss in MA or LA scenarios can be dramatical if the limiting factors (multipath, NLoS) occur. For indoor scenarios usually no GNSS based positioning is possible due to too weak satellite signals. Of course, the additional receiver hardware leads to higher power consumption.

For a general solution a hybrid approach is suitable depending on the UT position and the environmental conditions. Usually, as much as possible of available information sources should be used for positioning. In WA scenarios where good LoS access to the satellites is possible, a GNSS based solution is the best choice with supporting information and measurements from the WINNER RAN. In LA or indoor scenarios, where no satellite signals are available, a pure WINNER RAN based location determination is necessary. In MA, it could be a WINNER RAN based solution where – if available - GNSSs signals are used to improve the positioning of the UT. Again, the limiting factors in these scenarios are determined by NLoS and multipath effects.

Usually, we differentiate between the following localization classifications [STK05] depending on the system component where the main parts of computation are realized:

1) Network based
In network based positioning all required measurements are performed by the network or BSs. Main advantage of this method is, that no changes at the handsets are necessary and therefore legacy UTs can be used. Nevertheless, the load of the network for this method can increase due to signalling operations of the BSs. The estimate of the UT is determined by the network and can directly be used for system specific applications. For user specific applications the final position estimate has to be retransmitted to the UT.

2) UT based
If the UT processes the measurements and determines finally the estimate of the UT position the method is called UT based positioning. For UT based processing changes on the handsets are required. The signals from several BSs are measured at the UT and postprocessed to determine the location. Advantage of this method is, that all UTs can use the same signals from the BSs for positioning without increasing network load. Nevertheless, the computational complexity of the UT for performing the location determination is increased due to the effort in terms of calculation power and equipment loading.

3) UT assisted
The UT assisted method is a combination between network based and UT based positioning. The terminal measures the available signals itself and afterwards, the measured data is retransmitted to the network. Finally, this data is postprocessed in the network for location determination. Clearly, the network load is the highest of the proposed methods but nevertheless, the computational complexity at the handset can be reduced compared to pure UT based processing.

For location determination by cellular networks using WINNER RAN each of the aforementioned procedures can be used. If we include GNSS systems in our investigations, a pure network based solution is not possible because the measured satellite signals have to be preprocessed directly at the UT.

In the following, we describe the measurements provided by the network or UT (cf. Figure 4-3) that can be used as input for the final location determination [GG05]. Additional to the pure measurements the variances should be provided by the entities as input for the further processing steps.

1) Time of Arrival (ToA)
In ToA the propagation time from the BS to the UT or vice versa is measured. The propagation time is proportional to the distance between these two units. To determine ToAs, full synchronization between the BS and the UT is required. Usually, the UT clock is not perfectly synchronized, so its clock bias must be treated as a nuisance parameter. Note, that the GNSS position solutions are based on ToA measurements. ToA can also be performed by round trip time measurements. For a pure ToA solution at least three independent ToAs are required for an unambiguous positioning in two dimensions. From a geometric point of view, the UT lies on the point of intersection of at least three circles around the BSs for two-dimensional processing.

2) Time Difference of Arrival (TDoA)
To avoid the effort of full synchronization between BSs and UT, the method TDoA can be used, i.e., by taking time differences of ToA measurements the clock bias nuisance parameter is eliminated. Therefore, only relative propagation times are measured relative to one arbitrary chosen reference BS. For this method, only full synchronization between the BSs is required. From a geometric point of view, the UT lies on the point of intersection of at least two hyperbolas (according to three BSs) with foci at the two related BSs for two-dimensional processing.

3) Received Signal Strength (RSS)
Because the transmit power is known and the received power can be measured by the system, the channel attenuation on the BS-UT link can be computed. This is directly related to the distance between the UT and the BS assuming adequate pathloss models [WIND54]. The geometric interpretation is similar to ToA.

4) Angle of Arrival (AoA)
Usually, this method is a pure uplink approach, where the angle of the incident wave of the signal transmitted from UT to BS is measured. Multiple antennas are required which is assumed only at the BSs. A simple method with restricted performance for very coarse estimates is based on sectorized antennas, e.g. antennas with three sectors to cover cell sectors of 120° [WIND210]. For AoA method only two BSs are required for an unambiguous solution in two-dimensional positioning.

5) Cell ID
The identification of the serving cell is always available and can give a very coarse estimation of the UT position.

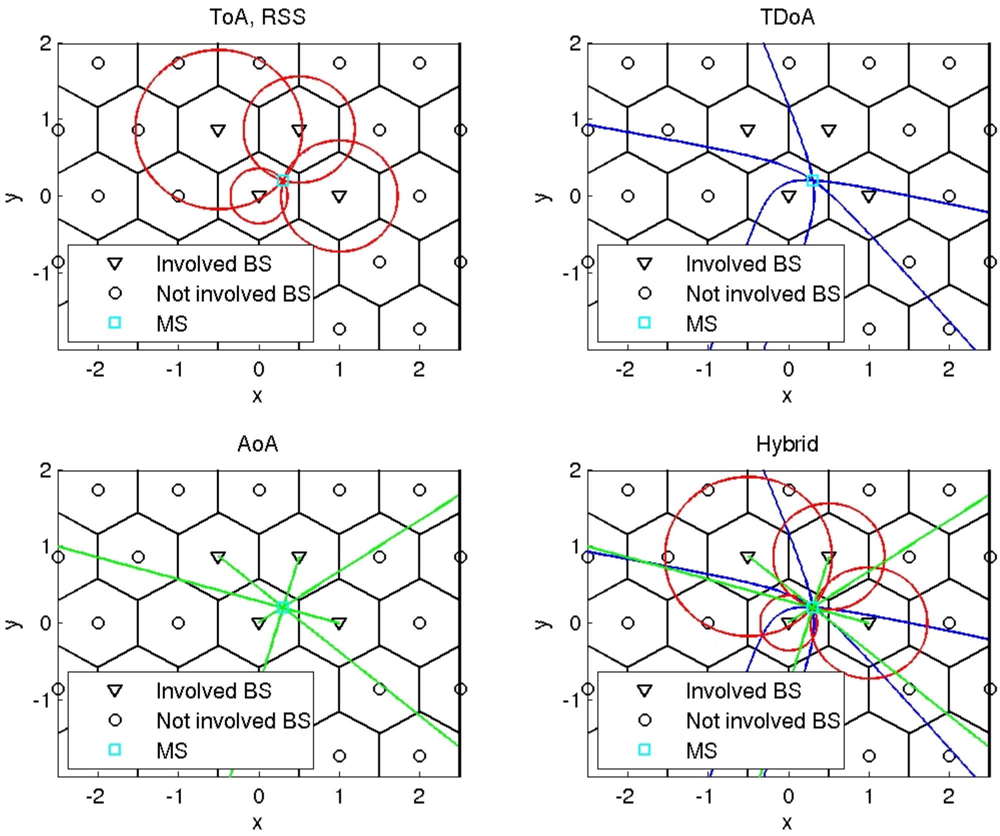

Figure 4-3: Measurements needed for location determination, two-dimensional processing based on four involved BSs

The accuracy of ToA and TDoA timing measurements can be very high and depend mainly on the bandwidth of the signal and the SNR, where SNR includes receiver noise, intra-, and intercell interference. RSS measurements depend strongly on the used pathloss models [WIND54] and the quality of SNR estimation. However, compared to the timing measurements (ToA, TDoA) only restricted performance is to expect. AoA measurements depend mainly on the distance between UT and BS which might be a problem especially in WA scenarios. Furthermore, the hardware requirements (due to multiple as well as very good synchronized antennas) at the BS are very high for this method. For all of these procedures the limiting factors are multipath and NLoS effects that introduce a bias in the measurements which cannot be compensated by simply averaging over several measurements. Nevertheless, also for perfect conditions (no multipath, LoS) the performance of the most promising timing measurements (ToA, TDoA) depend mainly on the bandwidth of the signal and the SNR. For example in a ToA scenario, the achievable accuracy of the timing measurements $\hat{\tau}$ is bounded by

$$\mathrm{var}(\hat{\tau}) \geq \frac{1}{SNR \cdot \overline{F}^2},$$

where $\overline{F}^2$ is the mean square bandwidth of the signal [Kay93]. Note, that the SNR can be increased by averaging over several measurements and hence, the timing measurement performance can be improved.

Related to network or UT based positioning is the question, if the location determination process for the WINNER RAN measurements should be based on uplink or downlink measurements. For certain parameters (e.g. AoA) only uplink measurements are reasonable, because we assume the necessary multiple antennas only at the BSs. On the other hand, e.g., TDoA measurements can theoretically be provided by uplink or downlink processing, depending on available resources on UT and RAN side. Note, that relevant parameters (e.g. bandwidth, OFDM structure) can differ for uplink and downlink measurements [WIND210] and therefore the performance can be different.

Based on all available measurements, the position information will be calculated as hard and as soft output, i.e., besides the pure coordinates of the position estimate also the variance or probability density function of the estimate will be calculated according to the soft location principle [AKR+01]. Furthermore, the direction of the UT and its velocity will be computed. The accuracy of the UT positioning estimate can be further improved if additional information about the history of the recent estimates and a suitable mobility model is involved. This allows to track the user on its way through the RAN. All of the delivered estimates can be used as input for the already mentioned user or system specific applications. An overview of the location determination process can be found in Figure 4-4.

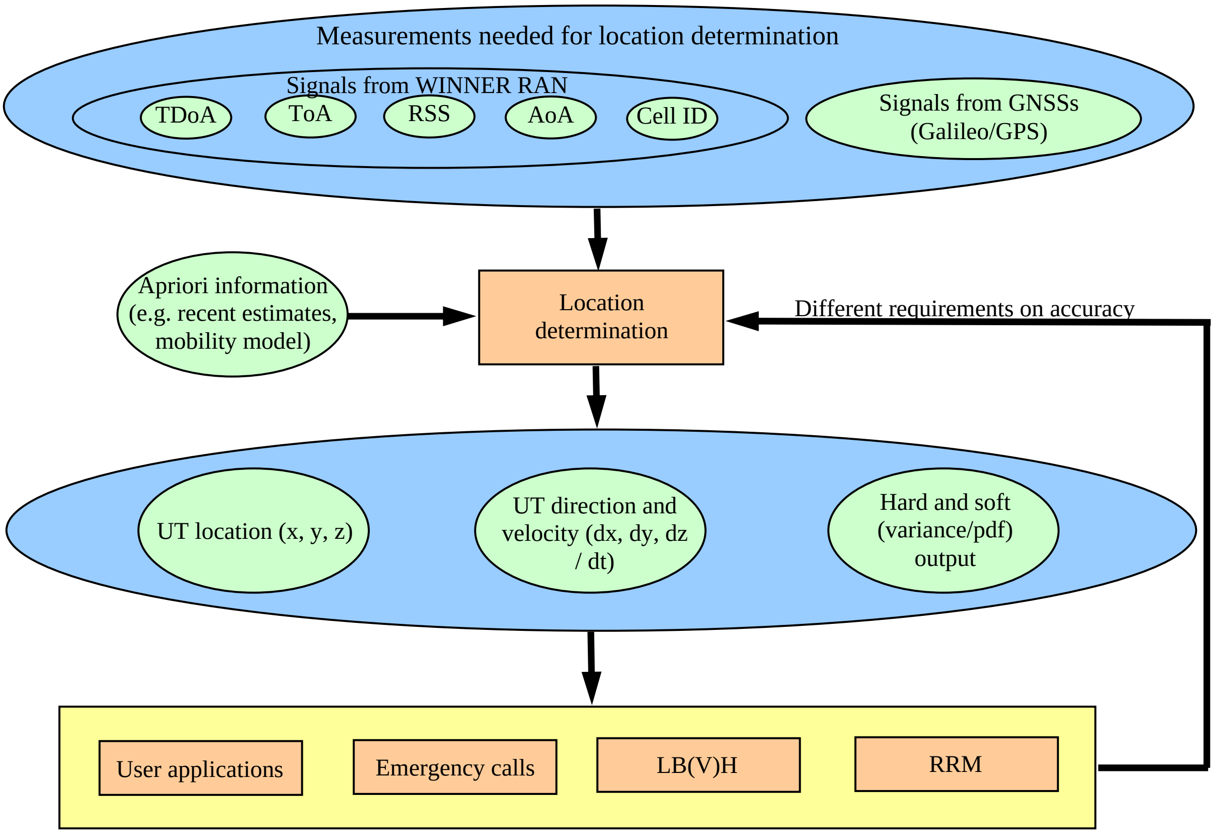

Figure 4-4: Location determination processing

In the following we point out the limiting effects for positioning based on timing measurements taken from the cellular network:

If multipath is present in a GNSS link, i.e., not only the LoS signal but at least one reflected and delayed version of the signal appears at the receiver, the signals superpose and yield a bias in the delay/pseudo-range estimation process. Of course, several effects have impact on the relevance and quantity of multipath effects in the UT, e.g., the bandwidth of the system. To combat this effect to achieve a sufficient performance, multipath detection and mitigation is a necessary procedure especially in critical environments like urban canyons in MA.

The NLoS problem is usually for communication systems not that critical as it is for navigation approaches. For communication issues it is not such a problem if the transmitted signal arrives at the receiver with a certain small delay caused by the non-direct path. But in a navigation scenario this means, that the NLoS path is erroneously identified as the direct path, and thus, the performance degrading bias is introduced in the timing measurements that can not simply be resolved by, e.g., averaging operations. Detection and mitigation of NLoS connections between the UT and the BSs is a difficult problem in navigation environments. In contrast to multipath mitigation algorithms that work on physical layer level, NLoS detection and mitigation is usually done by higher layer processing of the received measurements.

Finally, in Table 4-1 the requirements for the location determination of emergency calls are shown. It is based on the US Federal Communications Commission (FCC) that states that all wireless service providers have to deliver the location of all emergency callers with specified accuracy [Zha02]. Note, that accuracy assumptions are not yet well defined by a common European agreement. The requirements are expressed in terms of Circular Error Probability (CEP).

Specification	Network based	UT based
CEP67	100m	50m
CEP95	300m	150m

Table 4-1: CEP specifications for the location determination of emergency calls

It is differentiated between network and MS based positioning and – for instance – CEP67 = 100m means, that at least 67% of the radial positioning errors should be smaller than 100m.

5. WINNER RRM architecture

For efficient management of the resources, the location of the RRM functions within the network architecture is an essential issue as it can affect the performance of the network due to extensive signalling and delays.

In WINNER phase I it was studied different architecture options, from centralized to distributed, with several variants most appropriate architecture to support, the RRM cooperation functionalities with the legacy RANs and for the WINNER intramode and intermode RRM functionalities.

- The most appropriated architecture option for cooperation with the legacy RANs was a centralized architecture with a central node, called CoopRRM that took the cooperation functionalities decisions. One advantages of this approach is that not too many changes should be introduced in the legacies because the complexity and functionality is moved to the CoopRRM, without introduce to much delay in sensitive procedures (HO, channel switching, call setup). When information about the amount of spectrum resources in other modes is available, algorithms perform better and the usage efficiency is increased. This also applies when considering not only surrounding cells but also different radio networks operating simultaneously, i.e. JRRM approach. Benefits of centralised RRM are achieved at the expense of a higher computational complexity since a larger interchange of information among network agents, thus increasing the signalling. Delay in signalling is higher than is distributed approach, but the reaction time is not as critical due the vertical handover and other functionalities are inherently not as fast as WINNER inter/intramode functionalities.

- Regarding the WINNER intermode and intramode RRM functionalities the centralized approach was the initial approach but distributed architecture was judged as the most convenient approach in this time critical functionalities as fast intermode/intramode handover, but it lead to a more complex Base Stations (BSwa / BSma).

In any case, there will be different decision levels and time scales of the same RRM functionality (for example inter-RAN, inter-mode and intra-mode handover), located in different nodes.

In UMTS also there are different levels of decision because most of the RRM functionalities are located in the Radio Network Controller (RNC), but terminals and Node B also contribute to parts of RRM, in terms of power and load control.

5.1 CoopRRM, SRRMW and SRRML

One of the requirements of the WINNER architecture is the interworking with the legacies RANs, this approach present the advantages over former wireless systems to allow the seamless introduction of the WINNER new air-interface exploiting the installed base of legacy systems. To achieve this goal is necessary to coordinate individual radio resources management (RRM), associated to each radio access network RAN. This architecture should be enough flexible to accommodate the current RANs and also the new systems that could be deployed.

Figure 5-1, Figure 5-2 and Figure 5-3 show the CoopRRM and SRRM interactions:
The CoopRRM, the entity in charge of the coordination of the inter-RAN RRM functionalities of WINNER and the legacy RANs, is foreseen to be located outside of WINNER RAN, occupying a neutral position amongst wireless networks.
The SRRM, the entity in charge to adapt the RANs to the cooperation with the CoopRRM. The SRRM in the legacy RANs (SRRML) will have two types of functionalities and interfaces with the legacy RANs, one for traffic monitoring and reporting of physical legacy nodes and other devoted to the direct actuation in the legacy RAN nodes, basically it will translate CoopRRM commands to the legacy RAN. The SRRM in WINNER RAN

(SRRMW) will have the monitoring and actuation functionalities and also the support of the WINNER RRM functionalities related with the legacy RANs and some RRM intermode coordination functionality.

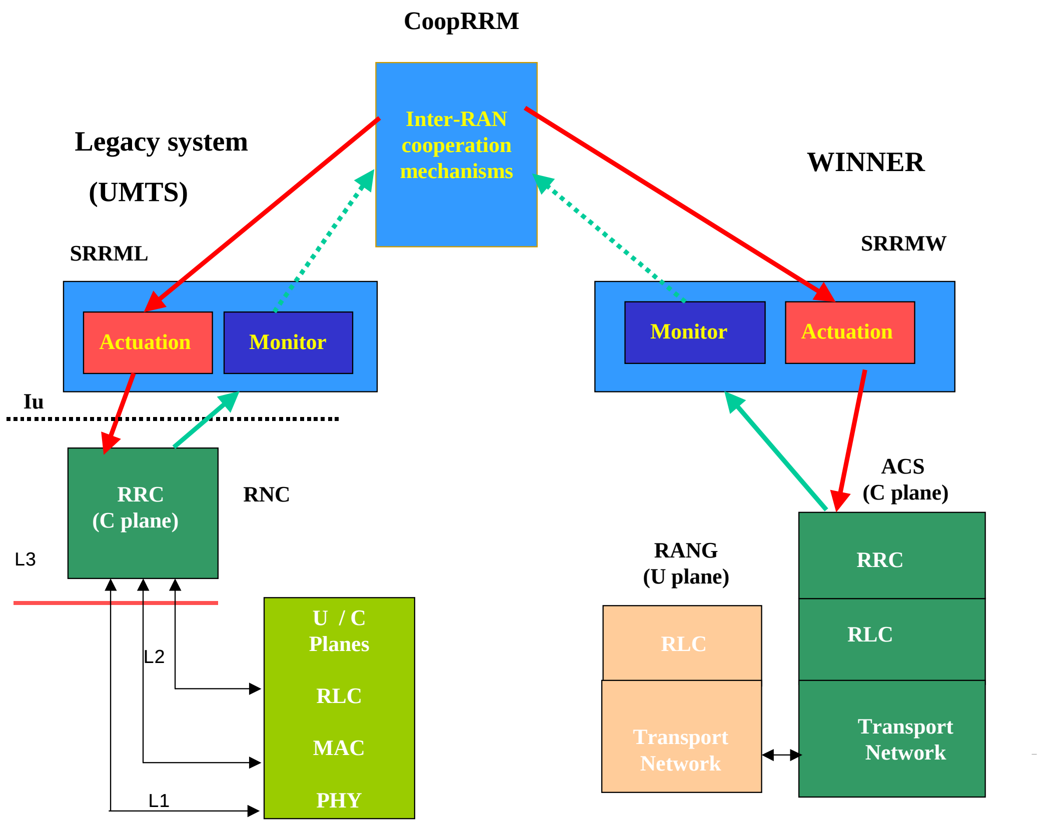

Figure 5-1: Logical nodes involved in the collaboration between WINNER RAN and a legacy RAN, including its layered architecture. UMTS is used as exemplary legacy network

There are two possibilities for the inter-RAN HO and other RRM mechanisms decisions, the CoopRRM only advises the SRRM entities and the SRRM are the masters or the CoopRRM decisions are binding for the SRRM entities.

The CoopRRM will have interfaces with other CoopRRM of the same or different operators.

The inter-RAN cooperation may be realized at lower layers, in consequence the SRRM will be associated to or reside in the RNC, BTS and MG of UMTS, GSM and IEEE802.11 networks, respectively.

5.2 WINNER RRM functions classification

The RRM functions could be classified within the three following categories:

1) ***Mode -specific RRM functions***: These are targeted and optimised for a specific mode and deployment scenario in terms of using different parameters specific to the current mode used. In the MAC system layer, they include scheduling, power control, and link adaptation. In the RLC control plane, they include intra-mode handover and routing.

2) ***Mode Generic RRM functions***: These are shared between the different WINNER modes or used for their coordination and include spectrum assignment, service level control, mode selection, buffer management, traffic policing, admission control, congestion control and inter-mode handover.

3) ***Cooperative RRM functions***, which are used for the cooperation of the WINNER system with legacy RANs such as UMTS and WLAN and reside in the Cooperative RRM (CoopRRM) entity. These include spectrum sharing, inter-system handover, admission control, congestion control and RAN selection.

5.2.1 Centralized architecture for the cooperation of WINNER modes

Currently UMTS and GSM are almost totally centralized systems in which only PHY layer functionality is associated to the BS (e.g. link adaptation) but new system needs to move user plane and control plane functionality towards the BS to solve the high delay inherent to the current mobile systems and provide cost-effective systems.

Figure 5-2 shows one implementation of the RRM architecture, basically followed in phase I.
According to this approach, the functionalities with higher time scales, the mode generic (user and control plane functions) are located at the RANG and ACS respectively, and the only ones with the lower time scales, the mode specific, are located at the BS and RN.

Regarding the user plane the RANG, with the service level controller has the overall responsibility for adjusting inter-flow fairness, assuring the fulfilment of service level contract agreements and total delay constraints. It may perform by requesting resources from several BSs that may utilize different WINNER modes and may allocate resources belonging to several modes/access points. As buffer management and traffic policing could impose constrains on the service load controller, they should be located at the same node. After flow classification, the service level controller (SLC) will direct packets of a flow to one specific queue within a buffer residing in a BS or RN.
On the contrary, the resource schedulers (RS), and routing should always be located close to the BSs to minimize delays, this mean that the BS can obtain metrics of cell load. Furthermore, routing and resource scheduling should work closely together to perform cross-optimization between routes and resource assignments.

Concerning to the ACS logical node, all mode generic control plane and cooperative functionalities are located at the ACS logical node. In this almost centralised approach the problem of optimizing resources is better posed as it is done in a common fashion rather than regarding separately cells/modes/RANs. The main disadvantage of such an RRM architecture is the higher computational complexity since and a larger interchange of information leading to the possible overload of the system due to increased signalling directed to the ACS node leading to high delay values.

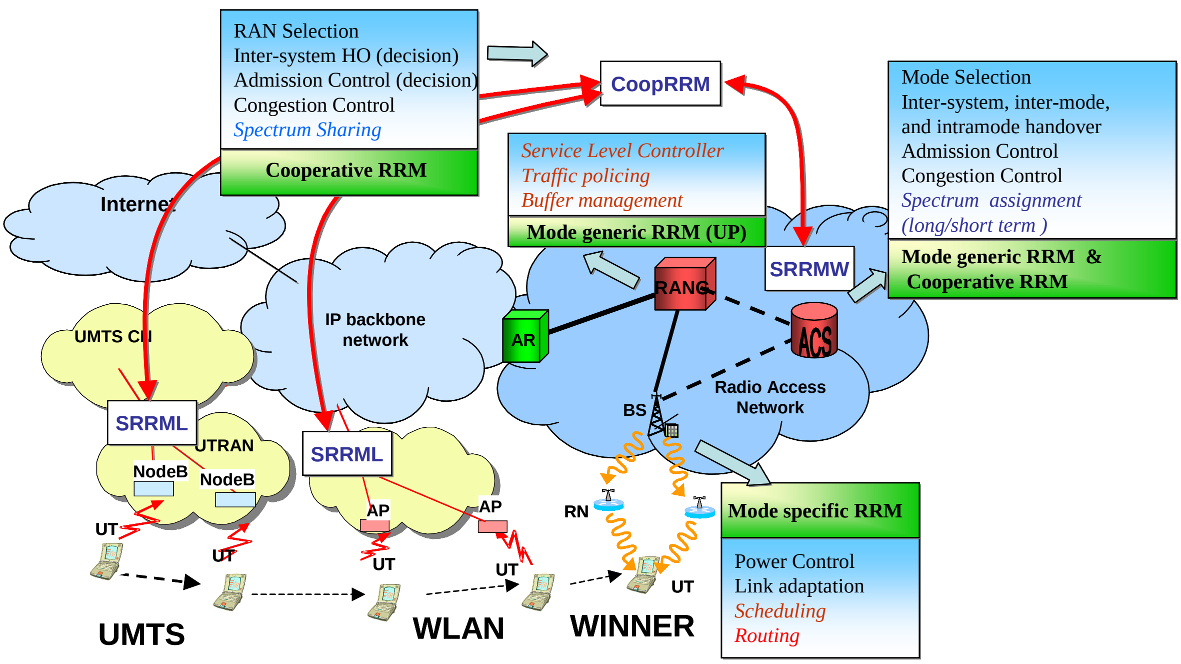

Figure 5-2: Mapping between RRM functionalities and logical nodes (WINNER modes with almost centralized architecture)

5.2.2 Partially Distributed Architecture for the cooperation of WINNER modes

Figure 5-3 presents another architecture option that could be useful to reduce the intermode RRM delays in decision making and signalling overload, in this option the BSwa and BSma will be more intelligent, with more RRM functionalities, moving a part of RRM functionalities from the ACS to the BS nodes. There will be an increase of the BS performance at RRM level but the cost will be an increase of complexity of the BSwa and BSma.

The ACS /SRRM would comprise mainly the RRM functions needed to the cooperation with the legacies (global WINNER Admission Control and Congestion Control, intersystem HO) and the coordination of the BSwa and BSma (when necessary) when a more comprehensive knowledge of the system is required. To allow inter-operability, the mode generic related control functions are located at the controlling BSs, e.g. BSwa or BSma. The mode generic functions can be admission control and handover control facilitated by the generic part of the interface between the controlled and controlling BSs. Due to the direct interworking between Bss, the amount of signalling information required will be much reduced which also allows faster decisions thanks to the local resources control at radio cell level.

For the BSwa/ma to be able to make decisions on mode selection / re-selection, inter-mode handover, admission control, congestion control and spectrum mapping between the BSla that fall within its cell as well as between it and the BSla, several information are required to be transmitted periodically or upon request from the BSla to the Bswa/ma. These include: BSla id and cell id, current load, max load, power information, handover statistics per RN and BSla (successful, drop rate etc), number of RNs attached to BSla (in case of mobile RNs), CELLla range (this can be change in case of mobile relays leaving/joining).
If we follow a self-organized and partially distributed approach the intramode handover decision could be taken by the BSs /UTs of the same mode, in a similar way to the current 802.11 standards, in which there is not a central entity over the BS as the current cellular systems. The BSs, of the same mode in the same deployment zone, could use a protocol to exchange control messages between them, in a similar way that the 802.11 APs use the IAPP protocol, to give a continuous coverage to support terminal mobility. In Figure 5-3, inter BSla is illustrated using the same colour as the interface between BSwa – BSla and BSma – BSla. However, due to the deployment differences between those cases, the inter-BSla might look physically different to the other case. There might be direct connection between BSla directly through the air, however, due to interference and power limitation, it is in the favour that the inter BSla connection is performed through the backbone.

Some of the mode specific RRM functions such as scheduling might reside at the RNs as it is envisioned to have RNs with different functionality layers (layers 1/2/3).

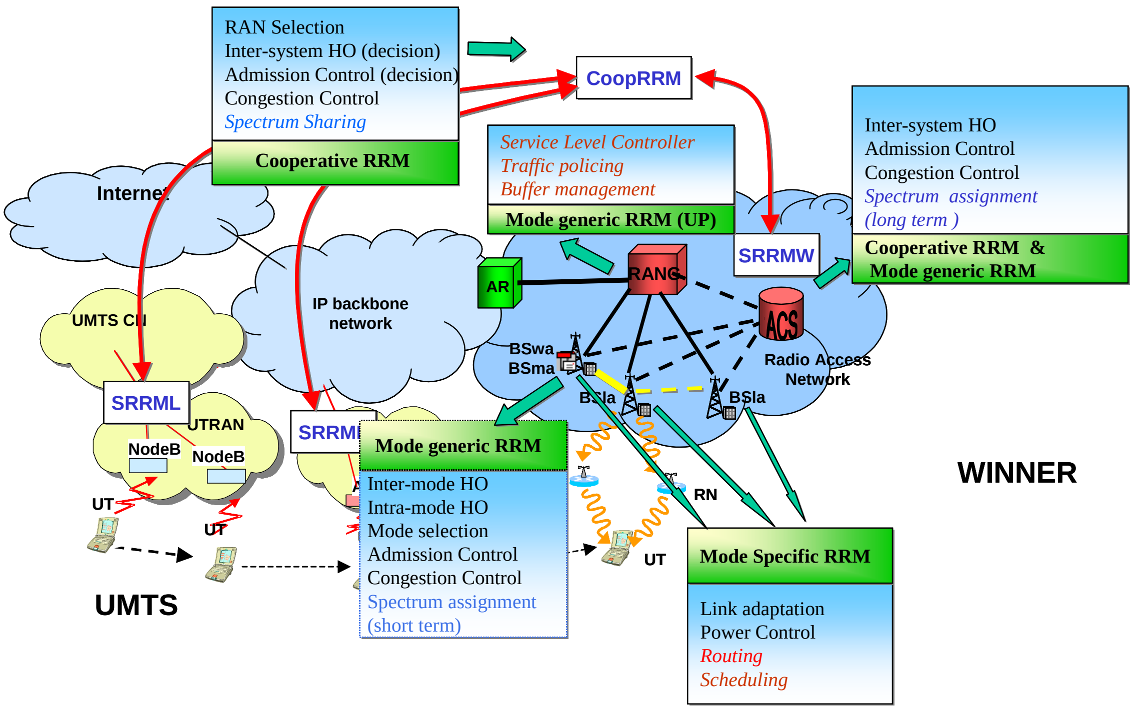

Figure 5-3: Mapping between RRM functionalities and logical nodes (WINNER modes with partially distributed architecture)

5.2.3 Inter-BS Transport and Direct Interworking

Assuming different types of BSs for the wide/metropolitan area and short range mode, we could restrict the extra functionality to the wide/metropolitan area 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 BSwa and BSma, making them responsible for the control and allocation of resources per wide area cell including all short range BSs (BSla) that fall within its coverage. A requirement for such an approach would be the definition of a communication link between the BSwa and BSla, denoted by a yellow line between them, 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 BSwa and Bsma and from other side their dependants BSla.

The interworking between BSs uses the air interface carried by the wireless spectrum resource as the deployment of the BSwa and BSla. The signalling link between BSwa and BSla is previously assigned. A possible interworking between the BSWA and BSLA is shown in section 2.2.1.2, where the control plane covering the BSwa, BSla as well the UT are illustrated by two hops by concept.

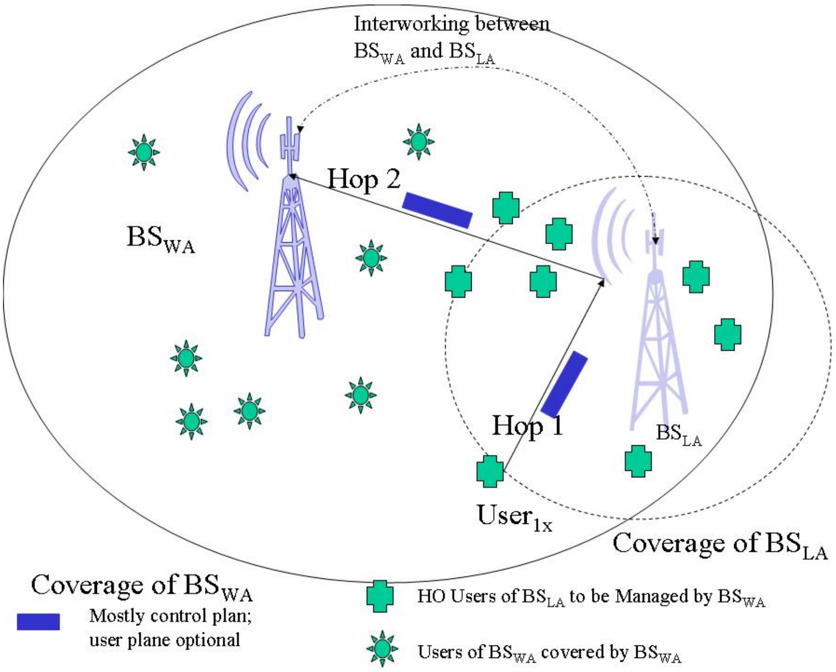

Figure 5-4: Illustration of interworking between BS_{WA} and BS_{LA}

In addition, all base stations and other network nodes can also be connected via an Ethernet based transport network. In generally the transport network should be opaque for the viewpoint of the protocol elements of the new broadband radio access network

6. Conclusion

This document described the cooperation mechanism between WINNER modes, considering the modes as a synonym for adaptation of the system to the application scenarios (wide area, metropolitan area and local area).
The cooperation mechanisms for the WINNER RAN that have been considered are the following: handover, congestion control (admission control and load control) and QoS based management.
Algorithms for intramode HO, based on signal strength and cell load, and using neighbouring cell list have been proposed, as well as algorithms for handover from wide area to local area that use as a decision criteria the parameters of the user terminal speed and cell load. Congestion avoidance control, encompasses Admission Control and Load Control with the objective the efficient management of radio resources. The Admission Control algorithm has to select the best cell for a new user/flow, but also the best mode to serve a specific user flow, considering several parameters. The Load Control algorithms aim at avoiding the network overload using preventive and reactive actions, using load sharing, forcing HO and decreasing QoS parameters in a coordinated way. QoS based management is related to the provision of a satisfactory to both users and providers QoS. The algorithm assumes that the different service classes use the most adequate mode, and considering different information profiles.
The cooperation mechanism for cooperation between WINNER and legacy RANs are, mobility management using fuzzy logic, (that could be extended to the cooperation of WINNER modes), congestion avoidance control, already proposed in phase I and here further extended and QoS based management that studies the needed KPIs, user profiles, user-oriented connection algorithms. High mobility support for the future multi-mode terminals is also discussed in this document, where cross-system broadcasting and joint radio resource management with performance comparison w.r.t. the terminal capabilities are provided.
In the user plane cooperation it is studied how to achieve high reliability in the user plane in a new scenario as is the integration of relays nodes in the MAC layer. Three approaches are discussed Layered ARQ, Relay ARQ and Multi-hop ARQ.
The WINNER RRM architecture has been considered separately for the two types of cooperation. For the support of the cooperation with the legacy RANs, a centralized approach is assumed, where the decisions are taken by the CoopRRM. For the support of the cooperation between the WINNER modes, two approaches have been presented: a centralized approach, in which the ACS controls the mode generic functionalities and a distributed (hierarchical) approach, in which the decisions are in charge of the BSwa that controls several RNs and the BSla. The trade-off between increase of performance and increase of complexity should be studied.
The Hybrid Information System (HIS) is proposed to be used for the improvement of the intermode handover increasing the cell size by precisely identifying the cell borders. HIS is a data base that contains the periodic UT and BS measurements, including its position.
Measurements are essential inputs for the wireless networks. It has been identified that the PHY, MAC and RLC measurements are necessary for the support of the proposed algorithms. Common pilots symbols in chunks will be used to evaluate link quality by received signal strength measurements. The inputs for the cell load computation have been identified; this information should be obtained by the Service Load Controller and from the MAC and the RRC control plane.
One of the features of the WINNER system is the UT location determination. One study of this issue is presented, determining the need measurements and its processing.

7. References

[AKR+01] M. Angermann, J. Kammann, P. Robertson, A. Steingass, and T. Strang, "Software Representation for Heterogenous Location Data Sources Using Probability Density Functions", Proc. International Symposium on Location Based Services for Cellular Users (LOCELLUS), Munich, Germany, February 2001.

[Ber01] P. Berzethy, J. Laiho and P. Szabadszallasi, "Dimensioning Process and Radio Resource Management in Third-Generation Mobile Networks", Nokia Networks, Third International Workshop of Design or Reliable Communication Networks, Budapest, Hungary, October 2001.

[Cod05] Pascal Codier, Didier Bourse, David Grandblaise, Klaus Moessner, Jijun Luo, Clemens Kloeck, et. al.: Cognitive Pilot Channel, Proccedings of WWRF15, Paris, 8.- 9.12.2005 CPC, WWRF

[GG05] F. Gustafsson and F. Gunnarsson, "Mobile Positioning Using Wireless Networks", IEEE Signal Processing Magazine, Vol. 22, No. 4, July 2005.

[Kay93] S. M. Kay, "Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory", Prentice Hall Signal Processing Series, 1993.

[Luo03] Jijun Luo, Markus Dillinger and Eiman Mohyeldin, Investigation on Radio Resource Scheduling in WLAN Coupled with 3G Cellular Network, IEEE Communication Magazine, June, 2003

[Luo04] J. Luo, E. Mohyeldin, M. Dillinger, etc., Performance Analysis of Joint Radio Resource Management for Reconfigurable Terminals with Multi-class Circuit-switched Services, WWRF, Toronto, Canada, 04~05. Nov., 2004

[Mih06] [A. Mihovska, et al., "Algorithms for QoS Management in Heterogeneous Environments," submitted to WPMC'06, to be held in September 2006, San Diego, California.]

[Moh06] E. Mohyeldin, J. Luo, etc., Common Pilot Method Enabling Network Assisted Fast Scanning for Reconfigurable Terminals, WSR 2005, Karlsruhe, 23-25th, March, 2006

[Mon96] E. Monteiro, G. Quadros, F. Boavida, "A scheme for the quantification of congestion in communication services and systems", IEEE Proceedings of Third International Workshop on Services in Distributed and Networked Environments, June 1996

[PS96] B. W. Parkinson and J. J. Spilker Jr., "Global Positioning System: Theory and Applications, Volume 1", Progress in Astronautics and Aeronautics, Volume 163, 1996.

[STK05] A. H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-Based Wireless Location", IEEE Signal Processing Magazine, Vol. 22, No. 4, July 2005.

[WIND210] WINNER, "D2.10: Final Report on Identified RI Key Technologies, System Concept, and Their Assessment", Deliverable, IST-2003-507581 WINNER, December 2005.

[WIND32] WINNER, "D3.2: Description of identified new relay based radio network deployment concepts and first assessment by comparison against benchmarks of well known deployment concepts using enhanced radio interface technologies", Internal Deliverable, IST-2003-507581 WINNER, February 2005.

[WIND35] WINNER, "D3.5: Proposal of the best Suited Deployment Concepts for the identified Scenarios and related RAN Protocols", Deliverable, IST-2003-507581 WINNER, December 2005.

[WIND41] WINNER, "D4.1: Identification and Definition of Cooperation Schemes Between RANs", Internal Deliverable, IST-2003-507581 WINNER, June 2004.

[WIND42] WINNER, "D4.2: Impact of Cooperation Schemes Between RANs", Internal Deliverable, IST-2003-507581 WINNER, February 2005.

[WIND43] WINNER, "D4.3: Identification, definition and assessment of cooperation schemes between RANs", Final deliverable", IST-2003-507581 WINNER, June 2005.

[WIND44] WINNER, "D4.4: Impact of Cooperation Schemes Between RANs", Final deliverable, IST-2003-507581 WINNER, November 2005.

[WIND45] WINNER, "D4.5: Assessment of cooperation mechanisms in system context", Final deliverable, IST-2003-507581 WINNER, December 2005.

[WIND54] WINNER, "D5.4: Final Report on Link Level and System Level Channel Models", Deliverable, IST-2003-507581 WINNER, September 2005.

[WIND76] WINNER, "D7.6: WINNER System Concept Design", Deliverable, IST-2003-507581 WINNER, October 2005.

[Zha02] Y. Zhao, "Standardization of Mobile Phone Positioning for 3G Systems", IEEE Communications Magazine, Vol. 40, No. 7, July 2002.