3GPP TR 25.881 V5.0.0 (2001-12)

(Proposed Technical Report)

3rd Generation Partnership Project;
Technical Specification Group Radio Access Network;
Improvement of RRM across RNS and RNS/BSS
(Release 5)

Keywords

<keyword[, keyword]>

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

http://www.3gpp.org

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

This document is for part of the Rel 5 study item "Improvement of RRM across RNS and RNS/BSS".

This study report describes the techniques to improve the RRM across RNS and RNS/BSS. It also deals with the feasibility of the techniques which are described in this document and their benefits.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.

[1] TR 101 112 V3.2.0, Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03 version 3.2.0)

[2] 3GPP TS 25.423: "RNSAP Specification".

[3] 3GPP TS 25.433: "NBAP Specification".

[4] 3GPP TR 21.905 V5.1.0, Vocabulary for 3GPP Specifications (Release 5)

[5] 3GPP TS 48.008, "3rd Generation Partnership Project; Technical Specification Group GSM EDGE Radio Access Network; Mobile-services Switching Centre - Base Station System (MSC - BSS) interface; Layer 3 specification"

[6] 3GPP TS 25.413, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UTRAN Iu interface RANAP signalling"

[7] 3GPP TR 25.935, "3rd Generation Partnership Project (3GPP); Technical Specification Group (TSG) RAN 3; Radio Resource Management (RRM); Optimisations for Iur and Iub"

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply.

CRRM controlled network: Describes the whole network which shares CRRM functionality. This network may consist of more than one RAN/RAT and may consider different cell layers (macro, micro, pico or wide area/local area).

MSS: Maximum Segment Size (used by TCP)

Multi-radio: more than one RAT or more than one radio access mode.

PS: Packet Switched

Radio access mode: different modes for the same RAT, e.g.: FDD and TDD for UTRA.

RTSP: Real Time Streaming Protocol

RTT: Round Trip Time, the time for a packet + response to travel, sender→ receiver and back

TCP: Transmission Control Protocol

Trunking gain: Increase in the overall capacity/throughput when sharing different pools of resources.

3.2 Symbols

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

CRMS	CRRM Server
CRRM	Common RRM
DR	Directed Retry, initial access cell change at call setup
IRNSAP	Inter Radio Network Subsystem Application Part, protocol on the Iur-g interface
Iur-g	Interface between RNC and 2G-BSC
NRT	Non Real Time
RAN	Radio Access Network, a system providing mobile access through coordination of a number of cells, e.g.: UTRAN, GERAN
RAT	Radio Access Technology, a type of technology used for radio access, e.g.: GSM, UTRA, cdma2000. One RAT can have different radio access modes.
RNS	Radio Network Subsystem
RRM	Radio Resource Management
RT	Real Time

4 Introduction

In 3GPP TSG RAN #11, it was decided to create Rel 5 SI, "Improvement of RRM across RNS and RNS/BSS" to investigate Multiradio Support study area.

In this document, we assume a future mobile network scenario characterised as follows:

- Different RAN may co-exist in the same area (e.g. UTRAN, GERAN, ..).
- Considering different RATs in every RAN multiple cells with radio access modes operating on different frequencies may overlap and multiple cell layers (macro, micro, pico or wide area/local area) will co-exist.
  It is assumed that every RAN will already have its own RRM for performances optimisation inside the RAN.
- Multi-mode mobiles can be connected to different RAN RAT/radio access modes/cell layers.

As a result, knowledge about each RAN/cell is needed to manage their resources and to optimise the overall system performance.

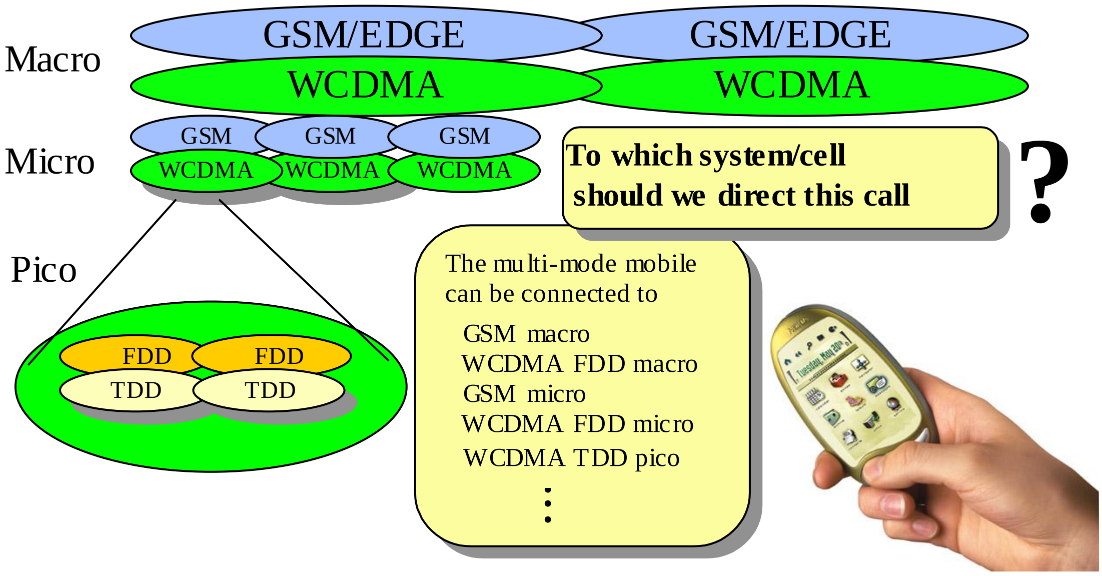


Figure 1. Future Mobile Network Environment

Such a management of resources is called Common Radio Resource Management (CRRM) and is characterized by the following functionalities:

- Coordination of different CRRM resource pools; one CRRM resource pool is characterized by having an own RRM functionality (so one pool may include one RAT or one radio access mode or one/more cell layer (macro, micro, pico or wide area/local area)/s or one/more operating frequency/ies);
- CRRM should fulfill the following task: Directing users in idle and connected mode to the cell and resource pool which is most suitable; what is most suitable may depend e.g. on the user's service (QoS requirements/classes) and network aspects like load balancing, minimizing interference, avoidance of needless handovers/cell changes etc.

To consider the above situation, it is necessary to investigate how to provide optimal resource management in a multi-radio environment and to optimise network performance.

5 Requirements

5.1 General

To provide optimal resource management and to optimise network performance in multi-radio environment, any solution(new/enhanced) should enable the followingperformance improvements:

Trunking gain: Less blocking from real time services. Higher throughput and lower delay for non real-time services. Lower probability of handover failure in case of load sharing /inter system handover.

Lower interference: Managing the average interference to be lower.

QoS Management: As a service can be realised in several systems, the call can be directed to the system which can best obtain the requested QoS with least cost for the operator.

5.2 Open Interfaces

If there is a need to define a new interface, then this interface shall be open:

Multi-vendor operation: An operator with for example GERAN equipment from vendor X and UTRAN equipment from vendor Y will have the possibility to steer and to control traffic between the systems.

5.3 Backward Compatibility

Backward compatibility: The solution shall be backward compatible and shall affect existing interfaces as little as possible.

Transparency to UE/MS: The solution for providing optimal resource management across RNS and RNS/BSS shall be transparent for UE/MS functionality.

6 Study Areas

Chapter 6.1 explains the necessary measurement and information to improve the RRM across RNS and RNS/BSS and chapter 6.2 shows the possible solutions, which has been discussed so far. Regardless the solution discussion, all solutions in chpater 6.2 are supposed to consider the measurement and the information, which is described in chapter 6.1.

6.1 Information needed for RRM Improvement

In order to improve the integrated CRRM functionality, the following functions have been identified as needed:

1. Measurements on Cells controlled by a distant entity (RNC or BSC) (i.e. dynamic information).

2. Static Information on Cells controlled by a distant entity (RNC or BSC).

6.1.1 Common measurements

The measurements such as load information of the neighbouring UMTS and GSM cells is an important input to the CRRM algorithm.

In Release 4, all the necessary procedures have been developed over the Iur to allow an RNC to obtain Cell Load measurements on Cells controlled by other RNCs.

The measurements between the BSC and RNC could be transferred in different ways; via Iur-g if existing, via the Iu and the A interfaces or via a new interface between any BSC and any RNC.

The following Common Measurements would be useful in an RNC for a decision on 3G to 2G handovers:

- 2G Cell load in CS Domain (e.g. Percentage of 2G Cell Capacity for CS Domain),
- 2G Cell load in PS Domain (e.g. Percentage of 2G Cell Capacity for PS Domain)

The following Common Measurement already available in RNSAP would be useful in a BSC for 2G to 3G handovers:

- 3G Cell load (e.g. Percentage of 3G Cell Capacity),

Possibly, UMTS-specific Common Measurement such as those already available in RNSAP could also be provided to the BSC:

- transmitted carrier power,
- received total wide band power,
- UL timeslot ISCP (for a TDD Cell).

6.1.2 Static Information on Cells controlled by distant entity (RNC or BSC)

6.1.2.1 Cell relation

The knowledge about the cell relations, i.e. overlapped, Macro/Micro cells or in general Hierarchical Cell Structure (HCS) is valuable in the SRNC/BSC in order to direct the traffic to the most optimal cell. Furthermore, the SRNC may not perform Inter-frequency CM measurements if the neighbouring cell is overlapped with the serving cell, and thereby avoiding capacity loss due to the CM measurements. In addition, the delays in the handover procedure will be reduced as the CM measurements are omitted.

Other HCS specific parameters such as HCS Prio (similar to HCS in RRC) may be required. Which parameters to include is FFS.

6.1.2.2 Cell capability

The GSM neighbouring cell information is normally setup via management interface in the CRNC. However, currently the SRNC does not receive all the necessary information over Iur about the neighbouring GSM cell capability. By the GSM cell capability, it is meant the information whether the GSM neighbouring is capable of, e.g. GPRS, HSCSD, EDGE, GERAN(Iu, Gb/A-interface), etc. This information is necessary for the CRRM algorithm in the SRNC in the cell candidate selection at call setup and handover.

6.1.2.3 Cell Capacity and QoS parameters

These two lists are preliminary and non-exhaustive.

The following information would be useful in an RNC for a decision on 3G to 2G handovers:

- 2G Cell Capacity for CS Domain (e.g. number of available Time Slots),
- 2G Cell Capacity for PS Domain (e.g. number of available Time Slots),
- Available QoS in PS Domain (e.g. Average Buffer Delay per Priority, Available Maximum Bit Rate per priority).

The following information already available in RNSAP would be useful in a BSC for 2G to 3G handovers:

- 3G Cell Capacity.

6.2 Proposed Solutions

6.2.1 CRRM Server

6.2.1.1 Introduction

This chapter introduces the concept of Common Radio Resource Management server (CRRM Server), as a new logical node in UTRAN and GERAN. The purpose of the CRRM Server is to collect resource management functionality that can be related to multi-system functionality, so that this can be processed in the same algorithm (i.e. handover, load control, QoS control, etc).

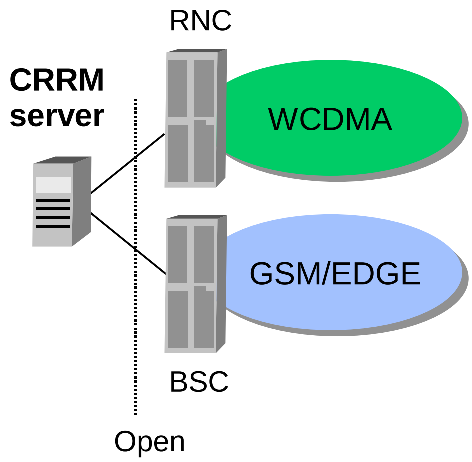


Figure 2. CRRM server relationship to the UTRAN/GERAN architecture.

With the introduction and integration of several systems with several modes and several layers, resource management becomes a more and more complicated task.

For example handover and load sharing algorithms shall not only maintain the connection at a reasonable quality, they should also consider whether it would be beneficial to move the connection to another system/layer/mode. This decision is not solely because of changing radio propagation, but also factors such as system load, operator priorities and service quality play an important role.

Although possibilities to gain information on other systems (e.g. inter system measurements GSM/UTRAN) are built in, the general understanding is that the respective algorithms shall execute in their respective system, and it is up to the operator to tune both systems to avoid ping-ponging, and to achieve optimal system performance.

In order to improve this situation, the introduction of a Common Radio Resource Management Server with an open interface towards both UTRAN RNS and GERAN BSS would make it possible to collect information from several systems/modes/layers and make a more optimal decision. CRRM server is not supposed to be be consulted for channel switching or for soft handover, since the RRM in RNC shall handle these kind cases. CRRM server will be used only for inter system/modes/layers HO.

6.2.1.2 Technical Details

Common Resource Management Server acts as a policy manager for the access to the cells and the radio bearer resources within UTRAN and GERAN, by performing the Radio Resource Management algorithms that are based on dynamic status information per cell from all the cells in the system. CRMS is also connected to other radio access network than UTRAN/GERAN in the future, allowing dynamic intersystem RRM.

One CRMS collects the information about the cell status in one area from RNSs or BSSs and, based on this, prioritises handover candidate lists.

6.2.1.2.1 CRMS Interfaces

The CRMS shall provide or ask for services to the logical entities(i.e. RNC, BSC and CRMS) that are connected to it. There are three main groups of services related to the CRMS:

1) Cell Measurement Gathering: This group of services is needed between CRMS and they are used to provide measurements related to the cells controlled by the RNC/BSC. This group of services may also be used by a CRMS to forward measurement to another CRMS (Edge Cell Measurement Gathering), which may be necessary in case the two CRMS control neighbouring cells.

2) Prioritisation of the list of candidate cells of a UE for a specific operation (handover, network controlled cell reselection...), This group of services is only needed between CRMS and RNC/BSC.

3) Error indication: this service is used by one entity to report detected errors in a received message, provided they cannot be reported by an appropriate response message. The error indication procedure can be used either between a CRMS and a RNC/BSC or between a CRMS and a CRMS.

These interfaces shall be standardised.

6.2.1.2.1.1 Cell Measurement Gathering

The measurement reporting procedures are based on the UTRAN Iub Common Measurement procedures [3] and/or on the Iur Common Measurement procedures [2]. Both of them, and also this interface, support on-demand (immediate), periodical or event-triggered reporting methods.

The CRMS shall select and set the reporting method for one, a group or all of the elements to be measured using just one requesting message. The reporting of the measurements shall be done in the same way as the request was done (individually or grouped)

Measurements can be expressed with parameters independent of the radio technology.

Four elementary procedures are defined for this purpose:

Elementary Procedure	Initiating Message	Response message	
		Successful Outcome	Unsuccessful Outcome
Measurement Initiation	MEASUREMENT INITIATION REQUEST	MEASUREMENT INITIATION RESPONSE	MEASUREMENT INITIATION FAILURE
Measurement Reporting	MEASUREMENT REPORT		
Measurement Termination	MEASUREMENT TERMINATION REQUEST		
Measurement Failure	MEASUREMENT FAILURE INDICATION		

6.2.1.2.1.2 Prioritisation of the list of candidate cells

The RNC/BSC sends to the CRMS the list of candidate cells of a UE for a specific operation (handover, cell change order...), including (when available) the mobile measurements for these cells and information about the quality of service that the user requires. The CRMS, after applying some algorithms, returns the prioritised list.

One 'class 1' elementary procedure will be defined for this interface.

Elementary Procedure	Initiating Message	Response message	
		Successful Outcome	Unsuccessful Outcome
Prioritised Cell List	PRIORITISED CELL LIST REQUEST	PRIORITISED CELL LIST RESPONSE	PRIORITISED CELL LIST FAILURE

6.2.1.2.1.3 Error Indication

The error indication procedure is initiated to report detected errors in a received message and same with the one defined in [2] or [3].

6.2.1.2.2 Example Scenario

6.2.1.2.2.1 UTRAN ⇔ GERAN

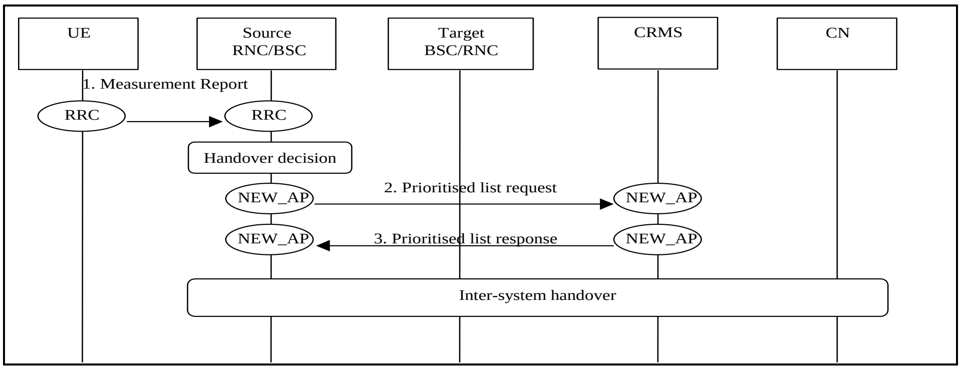


Figure 3. UTRAN ⇔ GERAN Intersystem HO with CRMS

1. Source RNC/BSC receives the neighbour cell measurements from UE and executes the handling for them.
2. The handover decision based on neighbour cell measurements is made. RNC/BSC sends the target cell proposal via Prioritised Cell List procedure in order to find the most optimal cell. Target cell list proposal is sent to the CRMS using the *Prioritised List Request* message..
3. The CRMS makes the prioritisation and returns the reordered candidate cell list using one *Prioritised List Response* message.

After target cell is decided, the procedure is same with the inter-system handover.

6.2.1.2.2.2 DRNC Covered by Another CRMS

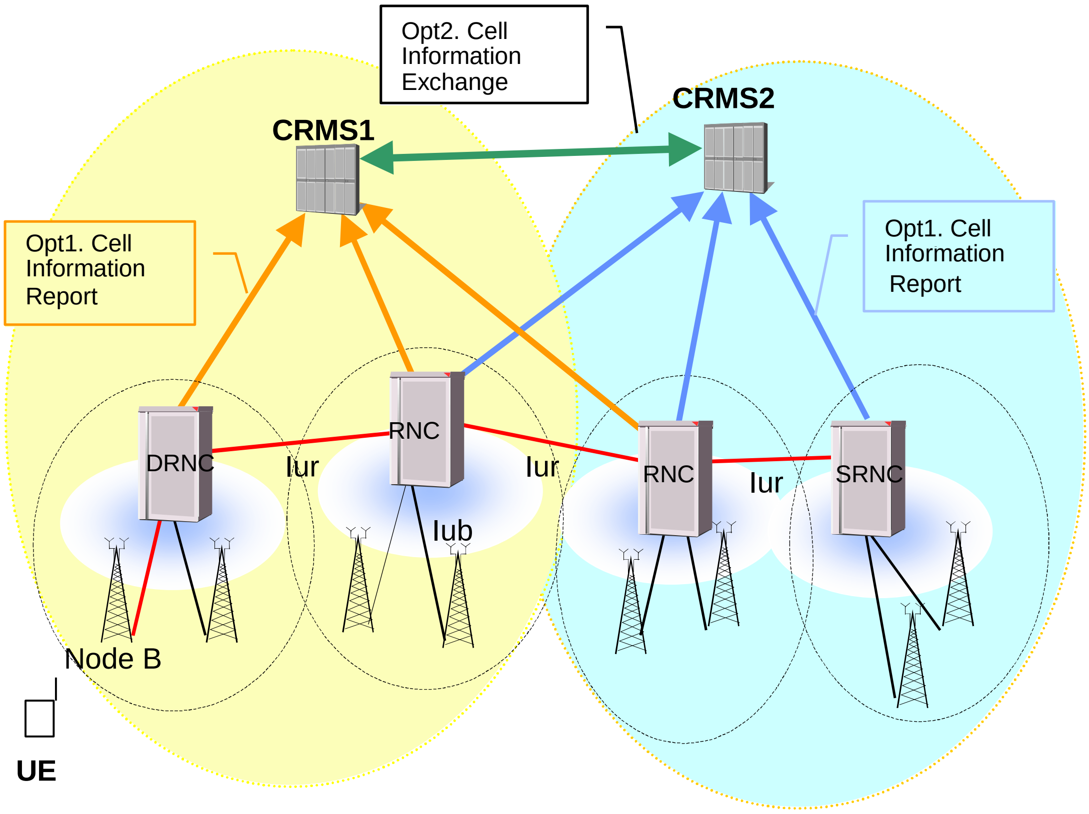


Figure 4. UTRAN Architecture with CRMS

Each CRMS will gather the cell load information under its coverage. For the neighbouring cells of the CRMS coverage, each CRMS can be provided the cell load information in 2 ways; Opt 1-CRMS can request cell load information of neighbouring cell or Opt 2-CRMSs can exchange the border cell load information.

It's up to each manufacture to choose which option to implement.

In case the DRNC is covered by another CRMS than the one covers SRNC as in [Figure 4], during RL Setup procedure DRNC will provide the address of CRMS1. Once SRNC decides HO and needs to consult CRMS covering the DRNC, SRNC will request the HO prioritised list from CRMS1.

6.2.1.2.2.3 UE in Soft Handover

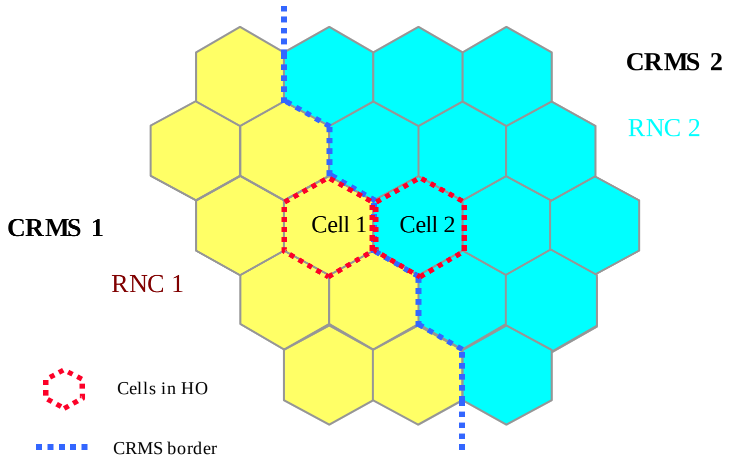


Figure 5. Border Cell SHO Case

If there are more than one CRMSs exist, then depending on the network planing each CRMS will keep the cell load information not only under its own coverage but also the neighbouring cells of CRMS border as described in 6.2.1.2.2.2. Therefore in case border cells are participating in SHO and each cell belongs to different CRMS coverage as [Figure 5], either CRMS can decide HO prioritised list because both CRMSs are having the border cell load information.

6.2.1.3 Open Issues

- CRMS-CRMS open interface. Is this interface needed ?
- The relationship between CRMS and RNC. In case there are DRNCs and more than one CRMSs, to which CRMS does SRNC ask the prioritized list (i.e.ordered list of candidate cells for operations like CRRM assisted inter resource pool handover)?
- If UE is in SHO and multiple CRMSs are involved in, which CRMS will decide prioritized list ?
- There is no mechanism to provide static information at the moment(It was not considered yet).

6.2.2 Integrated CRRM

6.2.2.1 Introduction

An alternative approach to achieve a CRRM functionality is to integrate it into the existing UTRAN nodes. As matter of fact many CRRM functionalities, such as Intra-RAT/Inter-frequency Handover, Directed retry, Service Handover, etc. are already supported in the current standard. The Iur and the proposed Iur-g interfaces already include almost all the required ingredients to support the CRRM functionality. Therefore, a natural approach is to continue this path and improve the existing CRRM functionality.

The main benefit of the Integrated CRRM is that with limited changes and already existing functionality it is possible to achieve optimal system performance. Most importantly, this is achieved without introducing additional delay which will deteriorate the delay sensitive procedures at call setup, handover and channel switching.

Furthermore, the additional delay will have adverse impact on the trunking gain especially for the bursty traffic which will cause reduced radio resource utilisation. In addition, delayed handover decision and execution will have negative impact on power control and thus reduced system capacity.

The delay requirement on the channel switching would in practice limit the possibility to interrogate the external CRMS, thus reducing the possibility to achieve a optimal system performance. Whereas, in the Integrated CRRM, the SRNC or BSC, based on its intra and inter system knowledge and the capacity, makes decision on whether to perform channel switching, inter-layer, inter-frequency or inter-RAT handover.

It is the general understanding that the RRM algorithms can not completely moved to an external CRRM. Some part of functionality will anyway reside in the RNS and BSS, which will result in that all three systems need to be tuned to achieve optimal performance, making the system tuning more cumbersome.

Different Integrated CRRM solutions are proposed that are more or less re-using the existing interfaces and the existing mechanisms.

6.2.2.2 Protocol and Architecture solutions

6.2.2.2.1 Common Measurements

For the BSC having an Iur-g, it is proposed to use Common Measurements over Iur-g.

For the BSCs that do not have the Iur-g, two solutions are proposed: Common Measurements over Iu/A interface or Common Measurements on a new interface between any BSC and any RNC (a sort of Iur-g light).

6.2.2.2.1.1 Common Measurements over Iur-g

Similar procedure as the RNSAP Common Measurement procedures are introduced.

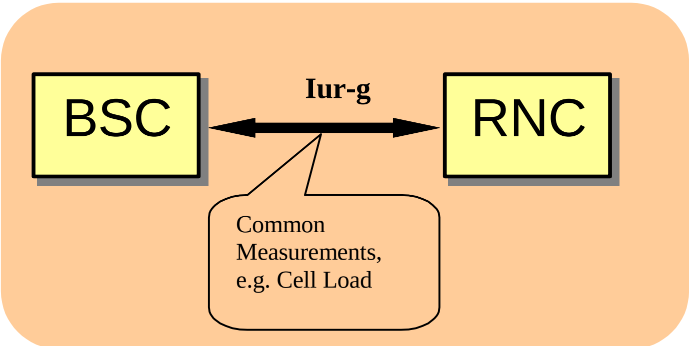


Figure 6. Common Measurements over Iur-g

6.2.2.2.1.2 Common Measurements when Iur-g does not exist

6.2.2.2.1.2.1 Common Measurements over Iu/A-interface

The cell load measurement is included in the transparent container IEs in the messages used by the RANAP Relocation procedures and BSSMAP Handover procedures.

Since the load measurements have short validity, this method is not as efficient as the Common Measurements over Iur-g. However, during high traffic the Inter-RAT handover is executed more often between the UTRAN and BSS which gives more frequent measurements which will result in more reliable load information. This method is an alternative especially if the Iur-g is not required for other reasons than cell load measurements.

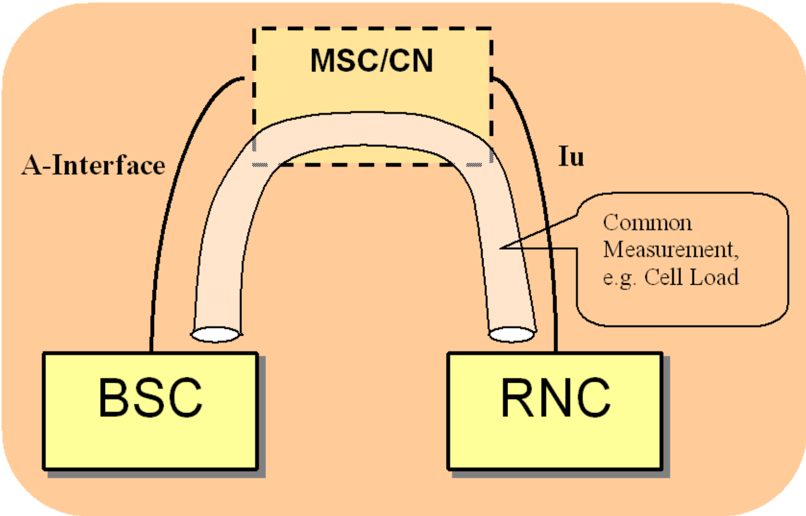


Figure 7. Common measurements in transparent containers over Iu/A-Interface

6.2.2.2.1.2.2 Common Measurements over Iur-g light

It is proposed to handle this by introducing Common Measurement procedures similar to the RNSAP Release 4 procedures in the Application Part of the new Iur-g light interface (for technical details on Iur-g light see § 6.2.2.2.3).

6.2.2.2.2 Static Information on Cells controlled by distant entity (RNC or BSC)

6.2.2.2.2.1 Solution using existing Iur features

The information about cell coverage area is proposed to be indicated to the SRNC by including the *Cell GAI* IE in the UMTS and GSM neighbouring cell information in the RADIO LINK SETUP RESPONSE and RADIO LINK ADDITION RESPONSE messages. The Cell GAI is already included in a number of the RNSAP messages from DRNC to SRNC, such as RADIO LINK SETUP RESPONSE for the UE cell-based positioning method.

It is further proposed to enhance the contents of the *GSM Neighbouring Cells Information* IE (in the RADIO LINK SETUP RESPONSE or RADIO LINK ADDITION RESPONSE messages) by also adding Cell Capability information (see § 6.1.2.2) as well as Cell Capacity and QoS Parameters (see § 6.1.2.3)

6.2.2.2.2.2 Solution using Iur-g or Iur-g light

It is proposed to handle this by introducing Information Exchange procedures similar to the RNSAP Release 4 procedures in the Application Part of the new Iur-g light interface (for technical details on Iur-g light see § 6.2.2.2.3).

6.2.2.2.3 Technical Details on Iur-g light solution

6.2.2.2.3.1 Introduction

It is proposed to define a new interface between an RNC and a BSC that allows to exchange common measurements and information on cells controlled by distant entities.

This interface should allow a connection towards a BSC supporting A/Gb mode i.e. it should not be linked to the support of GERAN Iu mode. Nevertheless, in order to anticipate the GERAN evolution in Iu mode and the interface called Iur-g, it is proposed that the interface between a RNC and a 2G BSC remains compatible with Iur-g. It means that this interface should be based on RNSAP. Furthermore the underlying transport should also be compatible with legacy 2G BSC equipments.

It is proposed to call this interface Iur-g light and IRNSAP (for Inter Radio Network Subsystem Application Part) the protocol on this interface.

6.2.2.2.3.2 Signalling Transport

Since the transport solution should be compatible with legacy 2G BSC equipments, B-SS7 cannot be used. Thus, it is proposed to use the IP stack that will be defined in Release 5 for the Iur interface.

6.2.2.2.3.3 IRNSAP functions

The following functions, taken from the R4 RNSAP protocol, are needed:

- Measurement on Common Resources.
- Information Exchange.

6.2.3 Extension of handover and relocation procedures

6.2.3.1 Principles

In addition to the solutions listed above for the exchange of cell traffic load measurements, another possibility is to reuse an existing procedure transferring information between BSCs/RNCs such as the Handover and Relocation procedures. The messages involved in these procedures can be easily extended to include the additional measurements that are required without the additional complexity of new nodes/interfaces and barely increasing interface load in the RAN/CN or processor load on the BSC/RNCs.

This section outlines the changes needed to the A and Iu interfaces to support this alternative solution.

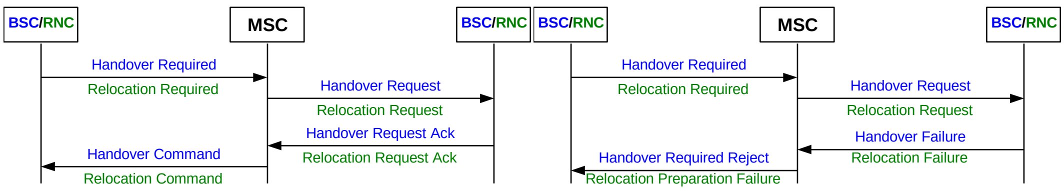


Figure 8 – Handover and Relocation procedures: successful and unsuccessful cases.

The successful and failure cases on the Handover and Relocation procedures are shown in Figure 8. Cell traffic load information can therefore be added to the following BSSMAP (see [5]) and RANAP (see [6]) messages for the 'forward direction':

- Handover Required / Relocation Required and
- Handover Request / Relocation Request

and to the following ones for the 'backward direction':

- Handover Request Acknowledgement / Relocation Request Acknowledgement,
- Handover Command / Relocation Command,
- Handover Failure / Relocation Failure and
- Handover Required Reject / Relocation Preparation Failure.

Note that this information can be defined in the same manner in both directions.

6.2.3.2 Use of this information by the BSCs/RNCs

6.2.3.2.1 Target BSC/RNC

When the target BSC/RNC receives a Handover Request / Relocation Request message with the cause value set to "*reduce load in the serving cell*", the target BSC/RNC compares the amount of free resources in the target cell with the amount of free resources in the serving cell (as indicated in the *Cell Traffic Load Measurement* IE), and if the target cell has more free capacity, it (typically) accepts the handover/relocation, else it (typically) rejects the handover/relocation.

For all Handover Request / Relocation Request messages (i.e. independently of the cause value), the target BSC/RNC stores the cell traffic load information for a period Tx. Tx is a parameter to help control the attempts by this BSC/RNC to transfer its own (excess) load towards that cell.

6.2.3.2.2 Source BSC/RNC

Irrespective of whether the handover or relocation procedure succeeds or fails, the cell traffic load information is sent from the target BSC/RNC to the serving one: for a time period Tx, this shall be taken into account for all future handovers or relocations from the serving node to the target cell.

6.2.3.3 Cell traffic load measurements

The format of the cell traffic load measurements needs to be standardised so that it can be used in a multivendor environment.

6.2.3.3.1 Load measurements

The minimum information to be included in the measurements is an indication of the use of the resources in the referred cell.

The simplest indication would be the relative amount of resources used/free. This, amongst other parameters, has been chosen in Release 4 for the common measurements over the Iur interface (see 3GPP TR 25.935 [7]). However, this does not allow a fair comparison due to the different capacity of each cell. The capacity of GERAN TDMA cells depends on the number of transceivers (TRXs). The capacity of UTRAN WCDMA cells also varies depending on the terrain, radio conditions, geographical distribution of the UEs, etc. For instance, a 6 TRX GERAN cell at 67% load should be favoured against a 2 TRX cell at 50% load.

Therefore, the best measurement to indicate is the **absolute** amount of **free capacity** (e.g. in Kbps) in the cell.

Although it may not be essential for Circuit Switched, the mentioned information should probably be separated into free uplink and downlink capacities. If the handover is accepted, the value indicated by the new BSC/RNC shall reflect the cell traffic load after the acceptance of the service.

6.2.3.3.2 Specific measurements

Specific measurements, other than the free capacity in the cell, may be reported. Some of them may be RAT specific.

6.2.3.4 Other considerations

In addition to the principles of the solution described above, there are other considerations to be taken into account:

a) **Unequal cell load information exchange**: Measurements from *popular* cells (i.e. those cells that are more likely to be handover candidates to/from a certain BSC/RNC) are reported more frequently, hence ensuring a more up to date load information in the BSC/RNC. Therefore, this solution optimises the compromise between the benefits of sharing cell traffic information and the additional signalling caused.

b) **"Sector" load**: when several UTRAN cells in different frequencies are collocated and covering the same geographical area, it should be specified that the RNC reports the sum of the free capacity in all those cells. It would be then the task of the RNC to ensure that the load is correctly distributed amongst the cells of the *sector*.

c) **Dynamic C31/C32**: cell traffic load information exchanged between BSCs can be used to update the C31/C32 GPRS parameters for cell reselection.

d) Validity of the measurements: as stated above, a timer may be needed (not necessarily in the standards) to govern the validity of the load information of neighbouring cells. The value of the timer is linked to the desired size of the interval of confidence of the cell traffic load Tx time units after the measurement is performed.

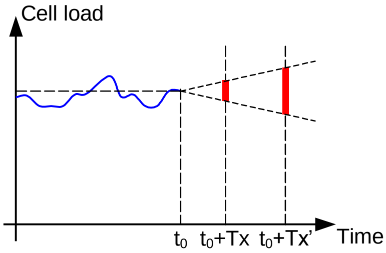


Figure 9 – Dependency of the interval of confidence size of the cell traffic load with the measurement validity timer.

e) Polling mechanism: if the validity timer expires, the BSC/RNC cannot use the last measurement received. An optional mechanism to poll the target BSC/RNC can be specified in order to get new traffic load information for a target cell. Two possible solutions are:

- Mechanism for the exchange of RAN information between BSCs/RNCs: this is currently being standardised by GERAN2 for Network Assisted Cell Change.
- *Dummy* handover/relocation: the source BSC/RNC may initiate an *erroneous* handover (i.e. a handover/relocation procedure resulting in a failure) in order to receive the traffic load measurement of the target cell on the Handover/Relocation Failure and Handover/ Relocation Preparation Failure messages.

f) *Compression* of traffic: it needs to be determined what the measurement should indicate in the case when the data rates of current MSs/UEs can be reduced: e.g. change from full rate to half rate, change of AMR mode, reduction of data rates of applications with background QoS class.

6.3 CRRM Performance Analysis

6.3.1 Simulation 1

6.3.1.1 Introduction

The target of this document is to quantify some of the benefits of Common Radio Resource Management (CRRM) for traffic management in an environment where several different radio access technologies co-exist with cells on several hierarchical layers.

One of the benefits comes from uniform distribution of traffic, which is desirable to maximise the trunking gain in network and to minimise the probability to make needless traffic reason handovers (see Figure 10). This also results in more uniform distribution of interference. This is achieved from the knowledge of the status of each cell, e.g. load in order to share the information of each cell.

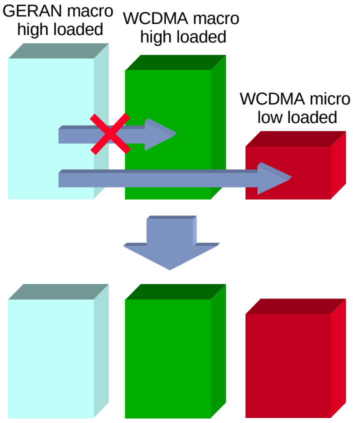


Figure 10. Trunking gain via load balancing

The trunking gain can be achieved by:

- directing a real time user to another system (or layer or frequency) by inter-system handover (IS-HO) or directed retry (DR) if a cell capacity threshold is exceeded - resulting in less blocking

- directing a non-real time user to another system (or layer or frequency) by inter-system network controlled cell reselection (IS-NCCRS) if the delay in the cell buffer is above a threshold – resulting in a higher average throughput (smaller average delay).

In this document the capacity (trunking) gains from CRRM concept are studied by relatively simple Matlab ® simulations. Only the load of the cells is considered as an input to cell prioritisation process. The simulations are carried out for both real-time (RT) and non-real-time (NRT) traffic. The simulation assumptions for RT traffic apply to both conversational and streaming traffic classes, whereas NRT assumptions apply only to interactive traffic class.

6.3.1.2 System Model

The simulation models are built on top of Matlab ® simulation tool. The simulation area consists of 5x4 hexagonal grid of cells. The border effects are alleviated by using wrap-around method (no UE enters or leaves the area). The left and right borders of the hexagon grid are connected to each other as well as the top to the bottom, thus each cell has 6 intra-layer neighbours.

There can be up to *n* layers on top of the first layer. Each layer has same properties, such as maximum bandwidth, i.e., radio system specific features/limitations (e.g. timeslots) are considered, a layer can be WCDMA, GSM, etc. No propagation model is modelled.

NOTE: The term ‘layer’ in this section 6.3.2 about the simulated system model is used synonymously to the term ‘resource pool’. So the term is not only related to the consideration of different cell layers (macro, micro, pico or wide area/local area) as in the rest of the TR. Also the term ‘inter system’ (e.g. in IS-HO and IS-NCCRS) used in 6.3.2 does not just refer to different RATs but in general to a change of resource pools.

In the simulations the position of a UE is not defined strictly, but as a membership to a certain cell. As with the call duration distributions these speed distributions are assumed from the information available. In Figure 11 the UE speed distribution used in the simulations is depicted.

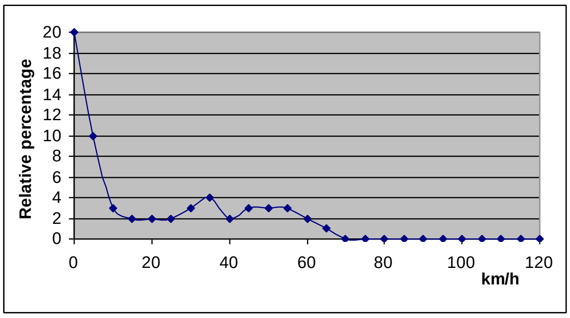


Figure 11. UE speed distribution

The direction of the movement is randomly determined. The border crossings between cells within the same system are treated as probabilities arising from the speed distribution of UE according to function

$$P_{MS\_BC} = \tanh\left[\frac{timestep \cdot UE\_Speed}{2 \cdot 3.6 \cdot L}\right], \qquad (1)$$

where *timestep* is one simulation step in [s], *UE_Speed* is speed of UE from the distribution in [km/h], and *L* equals to cell radius in [m].

The factor 3.6 in the denominator is a unit conversion from [km/h] to [m/s]. The factor 2 is needed as an adjustment, so that the border crossing probability is about 0.5, if *UE* moves one cell radius in one timestep. Tanh-function is used to make sure that the border crossing probability of a *UE* never exceeds 1, even if *timestep* or *UE_Speed* are set to a very high value.

6.3.1.2.1 RT Model and Algorithms

In this section the simulation assumptions for real-time simulations are described. RT simulations do not distinguish between conversational and streaming traffic classes.

6.3.1.2.1.1 Traffic Model

The air interface of the RT simulations consists of capacity definitions for each layer. In the model only hard capacity limits are used, such as maximum capacity for all cells in kbits/s. Thus, If the limits are exceeded, the UE is either handed over to another layer or the call is dropped.

For each single real-time call a certain duration time was chosen. The call duration distribution follows the formula

$$f(x) = \frac{1}{d}e^{-dx} \qquad (2)$$

where the mean duration of a call is 120s (d=1/120). All terminals are supposed to be multi-mode, so they can operate in all systems/layers. The traffic model is a constant bit rate model, with 100 % of activity, hence no silent periods are modeled.

The call arrival process has been modelled simply as

$$NumNewCalls = \frac{timestep}{average\ call\ length} \cdot Total\ number\ of\ active\ UE \qquad (3)$$

On the average the number of active calls remain constant during the simulation. Also with this approach it is assumed that no queuing occurs. Although for a real time services calls could be generated according to a Poisson process, it does not have much impact on the average capacity results.

6.3.1.2.1.2 Reference IS-HO/DR Algorithms without CRRM

Without CRRM only very little information about the target cell can be obtained. It is possible that there is no direct information available in one system about the load situation in the other system. So, an inter-system handover (IS-HO) attempt can fail due to high load in the target cell just not being known. In this case, the UE remains in its original cell. In these simulations the load reason HO triggering thresholds are located at 80% load for each cell. Above it the *UE*'s are handed over to another system because of load reason.

As the load reason IS-HO is triggered a number of users causing excess load are commanded to make inter-layer/system HO to randomly chosen target cells. If the target cell load is less than 80% the inter-system/layer HO is proceeded, but in contrary case ISHO is failed and call is kept in original cell. The procedure is illustrated in Figure 12 (left). If both current cell and target cell are fully loaded the call is dropped.

In initial access (call setup) same load thresholds are used as in case of handovers. If the load of the source cell is over 80% directed retry (DR) is performed to randomly chosen target cell. See Figure 12.

Note that only one DR is allowed in RT call setup and HO to randomly selected target cell among *n* layers.

6.3.1.2.1.3 CRRM IS-HO/DR Algorithms

Basically the same procedure is used in CRRM simulations as without CRRM. The only difference is that the cell capabilities are known by CRRM. Thus, the most optimum cell can be chosen among *n* parallel cells in both call setup and inter-system/layer HO. In Figure 12 (right) procedure with CRRM is illustrated. Note that the target cell prioritisation is based only on the cell load and no QoS issues are considered.

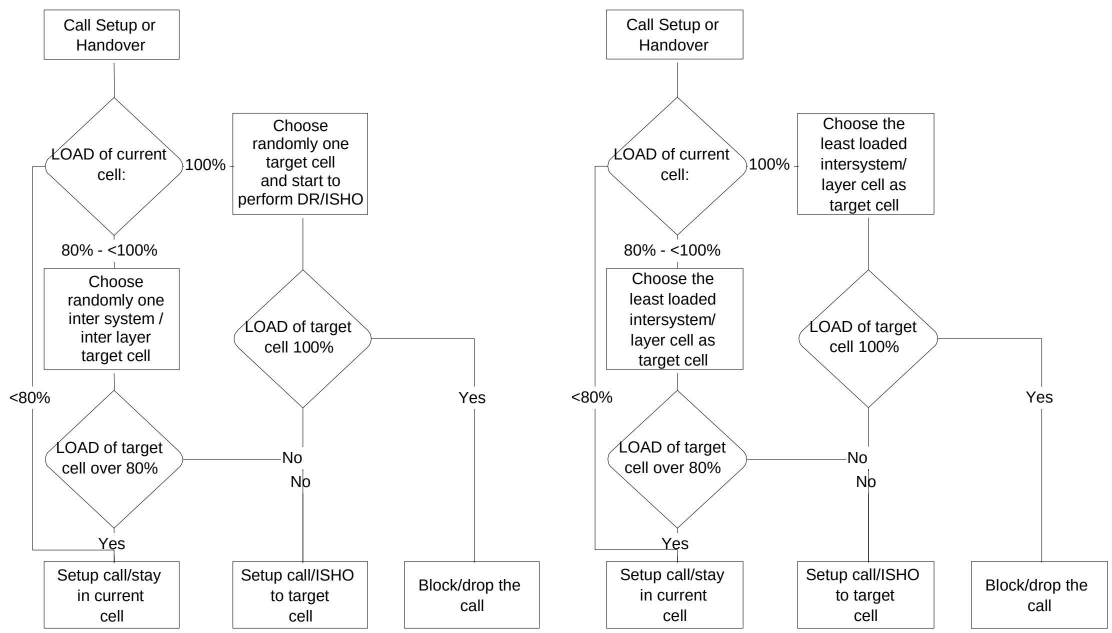


Figure 12. Call setup and traffic reason HO procedure without CRRM (left) and with CRRM (right)

6.3.1.2.2 NRT Model and Algorithms

In this section the simulation assumptions for non-real-time simulations are described. Here NRT simulations refer only to interactive traffic class.

6.3.1.2.2.1 Packet Traffic Model

In NRT simulations a typical WWW browsing model based on [1] is used. In the packet generating model one browsing session consists of a sequence of packet calls, and a packet call corresponds the downloading of a WWW document. After the document is entirely arrived to the terminal, the user is consuming certain amount of time for studying the information. This time interval is called **reading time**.

The user initiates a packet call when requesting an information entity. During a packet call several packets may be generated, which means that the packet call constitutes of a bursty sequence of packets (Figure 13). This phenomenon is not taken into account in the traffic model, since only the capacity or trunking gain results are in scope of these simulations. Therefore, the packet calls are not divided to smaller segments of data (packets), neither the TCP/IP rate adaptation mechanisms to include the packet arrival process within a packet call is included.

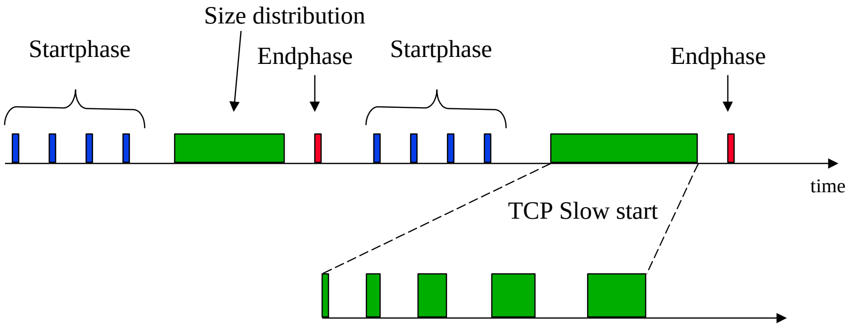


Figure 13. Generic model for NRT traffic.

Hence, only the following processes/variables are modelled in order to catch the typical behaviour of WWW-browsing (see Figure 14): Session arrival process, Number of packet calls per session ($\boldsymbol{N_{pc}}$), Reading time between packet calls ($\boldsymbol{D_{pc}}$), and Size of a packet call ($\boldsymbol{S_d}$). Note that the session length is modelled implicitly by the number of events during the session.

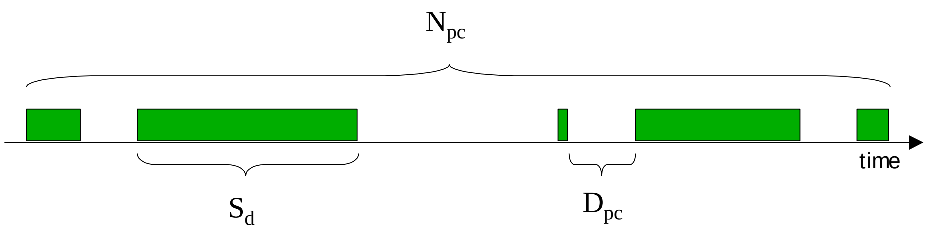


Figure 14. A packet session

Next it is described how these four different events are modelled. The geometrical distribution is used (discrete representation of the exponential distribution), since the simulations are using discrete time scale [1].

Session arrival process: The arrival of session set-ups to the network is modelled as a Poisson process. It is important to note that this process only generates the time instants when service calls begin and it has *nothing to do with call termination*.

The number of packet call requests per session, N_{pc}: This is a geometrically distributed random variable with a mean μ_{Npc} [packet calls], i.e.,

$$N_{pc} \in Geom(\mu_{Npc}).$$

The reading time between two consecutive packet call requests in a session, D_{pc}: : This is a geometrically distributed random variable with a mean μ_{Dpc} [model time steps], i.e.,

$$D_{pc} \in Geom(\mu_{Dpc}).$$

Note that the reading time starts when the last packet of the packet call is completely received by the user. The reading time ends when the user makes a request for the next packet call.

Packet call size, S_d: The traffic model can use such packet call size distribution that suits best for the traffic case under study. Pareto distribution is used.

The Pareto distribution is defined by:

$$
\begin{cases}
f_x(x) = \dfrac{\alpha \cdot k^{\alpha}}{x^{\alpha+1}}, x \ge k \\
F_x(x) = 1 - \left(\dfrac{k}{x}\right)^{\alpha}, x \ge k \\
\mu = \dfrac{k\alpha - m \cdot \left(\dfrac{k}{m}\right)^{\alpha}}{\alpha - 1}, \alpha > 1
\end{cases}
$$

In Table 1 default mean values for the distributions of typical www service are given. According to the values for α and k in the Pareto distribution, the average packet call size is set to 25 kbytes. The parameters of heavy-tailed Pareto distribution (packet call size) has been tailored to reduce the simulation time, thus packet call size can vary between 4.5 kbytes and 2Mbytes.

Process	Random Variable	Parameters
Packet Call Size	Pareto with cutoff	α=1.1, k=4.5Kbytes, m=2 Mbytes, μ=25Kbytes
Mean number of packet calls per session	Geometric	5
Mean reading time	Geometric	5 s

Table 1. Default mean values for the distributions of typical www service

6.3.1.2.2.2 Channel model and packet scheduling

The simulation setup is simplified by assuming each cell to have same capacity (bandwidth). Generated packet traffic is led through a time divided shared channel (e.g. DSCH, HSDPA) utilising the whole cell bandwidth. Each cell has a buffer where packet calls are scheduled in 'first in first out' basis (see Figure 15), and hence, each cell buffer is reduced by timestep*bandwidth kbytes in each simulation step.

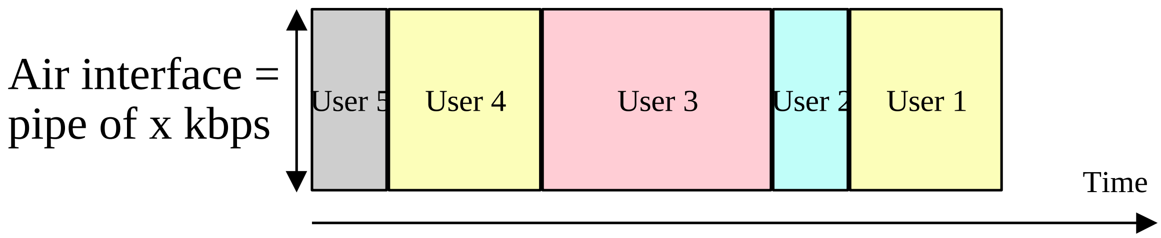


Figure 15. First in first out scheduling

6.3.1.2.2.3 Reference case

In reference case (no CRRM) simulations the inter-system network controlled cell reselection (IS-NCCRS) due to delay reason is not implemented for NRT data, and thus each layer/system generates and processes packet traffic separately.

6.3.1.2.2.4 CRRM IS-NCCRS algorithms for NRT simulations

In simulations with CRRM a delay reason Inter-system/layer network controlled cell reselection is triggered if the delay in cell buffer exceeds certain time in seconds (MaxDelayTreshold). CRRM checks if there is significant difference (MinDelayDifference) between current cell and the parallel IS cell having smallest buffer delay. The last packet calls in buffer are moved to parallel cell having smallest buffer delay such that the delay is balanced between cells. The purpose of these thresholds (MaxDelayTreshold and MinDelayDifference) is to reduce unnecessary cell reselection/signalling, and obviously there's trade-off between optimum performance and the amount of signalling.

Also in call setup a call can be directed to a cell where the buffer size is the smallest by directed retry. However, this is not very easy to implement, since in reality it would require measurements from other cells before call setup, and hence, cause much more delay.

Initial access cell change in call setup (directed retry DR) can also be understood so that the idle mode control of UEs is optimal, and new calls are always started in optimal layer.

6.3.1.2.3 Simulation Results

In this chapter the simulations results for both RT and NRT data are presented.

6.3.1.2.3.1 Real-time (conversational and streaming traffic) simulations

Main parameters for RT simulations are shown in Table 2.

Channel bandwidth	800 kbps
Service kbps	32 kbps, 144kbps and 384 kbps
Max connections per cell/layer	25 (32 kbps), 5 (144 kbps), 2 (384 kbps)
Load threshold	650 kbps (~80%)
ExamineMinutes (simulation time)	20-40 minutes

Table 2. RT simulation parameters

In Figure 16 two examples of blocking as a function of 144kbps user traffic without and with CRRM algorithms are shown. The number of users at 2% blocking is collected and plotted in Figure 18.

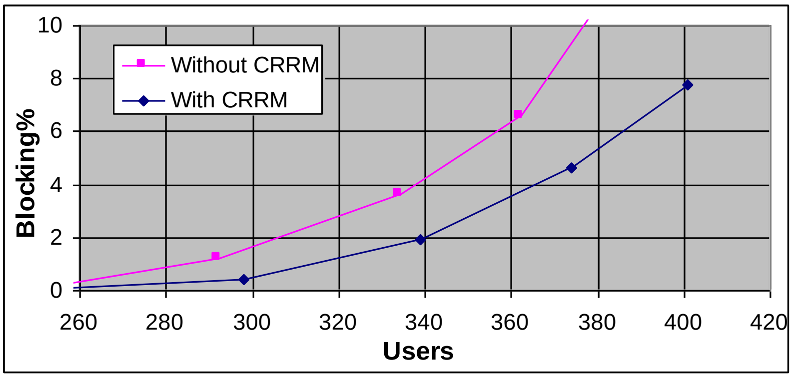


Figure 16. An example of blocking as a function of 144kbps user traffic without and with CRRM algorithms (5 layers)

In following Figure 17 – Figure 19 the system capacity and capacity improvements from different algorithms are plotted for different RT services. The abbreviations of simulation cases used in figures are explained in Table 3. In all cases the total number of users per each layer is equal on average.

Ideal ISHO/DR	Maximum capacity by ideal inter-system/layer handover and DR with all cells in same channel pool. Values obtained from Erlang B formula.
No ISHO/DR	Inter-system/layer handovers and DR not used. Traffic in each layer handled separately. Values obtained from Erlang B formula.
CRRM ISHO/DR	Inter-system/layer handovers and DR based on CRRM algorithms. See section 6.3.1.2.1.3for more details. Values obtained from simulations.
ISHO/DR	Reference Inter-system/layer handover and DR algorithm without CRRM. See section

without CRRM	6.3.1.2.1.2 for more details. Values obtained from simulations.

Table 3. The abbreviations of RT simulation cases.

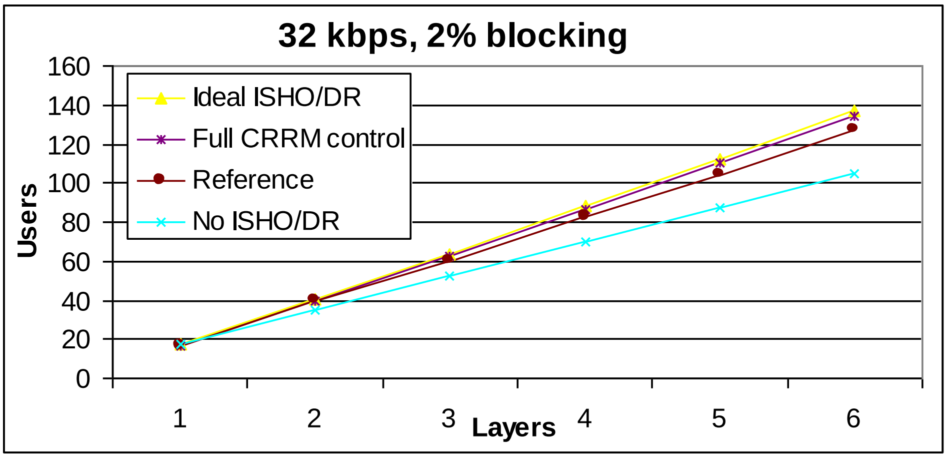


Figure 17. Average number of active 32 kbps users per hexagonal cell area on all layers

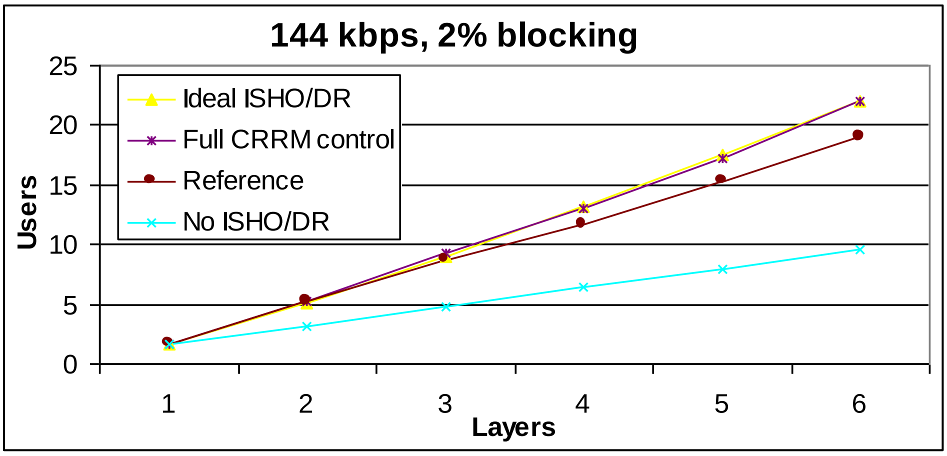


Figure 18. Average number of active 144 kbps users per hexagonal cell area on all layers

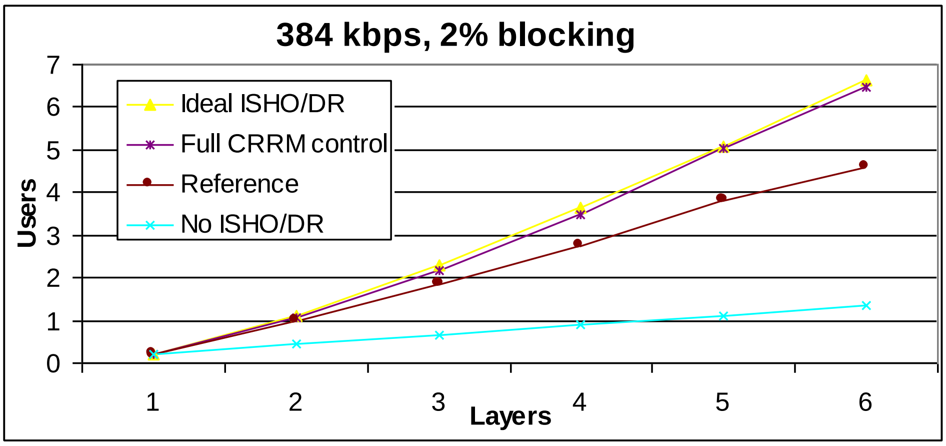


Figure 19. Active number of active 384 kbps users per hexagonal cell area on all layers

Previous figures can be summarised so that the higher the bit rate (less capacity), the more trunking gain can be obtained from CRRM when compared to the reference case without CRRM. For example the number of 384kbps users with 4 layers can be increased by 27% (11% for 144kbps, 4% for 32 kbps) with CRRM as compared to reference case.

6.3.1.2.3.2 Non-real-time (interactive traffic) simulations

Main parameters for NRT simulations are shown in Table 4.

Channel bandwidth	40 kBps (320 kbps) and 200 kBps (1.6 Mbps)
MaxDelayTreshold	5 seconds
MinDelayDifference	50%
ExamineMinutes	40-80 minutes

Table 4. NRT simulation parameters

The abbreviations of NRT simulation cases used in figures are explained in Table 5.

No IS-NCCRS/DR, unbalanced traffic	No cell reselections and no directed retry between layers/systems. Unbalanced idle mode, i.e. no CRRM: layer *n* has *n* times more users than layer 1.
No IS-NCCRS/DR, balanced traffic	No cell reselections between layers/systems. No directed retry between layers/systems. Users distributed equally between layers = CRRM balances users in idle mode
CRRM: DR only	CRRM directs new NRT users to the lowest loaded layer/system (least occupied cell buffer, i.e. smallest delay)
CRRM: IS-NCCRS only	CRRM moves users from the highest to the lowest loaded layer if load difference >50%. See section 6.3.1.2.2.4 for more details.
CRRM: DR and IS-NCCRS	Both initial access direction and load reason cell re-selection across layers/systems are used

Table 5. The abbreviations of NRT simulation cases.

6.3.1.2.3.3 Channel bandwidth 320 kbps

In Table 6 an example of results from CRRM simulations with both initial access and IS-NCCRS algorithms (6 layers) is shown. Correspondingly, cumulative distribution function of packet delay for the same case is plotted in Figure 20. The number of active users at 95% and 90% outage (delay less or equal than 5s) are collected (**bold**) and plotted in Figure 21 and Figure 22, respectively.

New calls/s	*Average actUEs*	*Users in buffer (ave)*	*Average buffer delay [s]*	*Number of packet sessions*	*Number of DRs*	*Number of packet calls*	*Delay reason IS-NCCRS*	*IS-NCCRS rate %*
20	599	184	3.6	46588	38278	232940	7964	3.42
23.33	**723**	236	4.2	54359	44760	271795	10895	4.01
26.67	875	316	5.1	62997	51922	314985	15296	4.86
30	**1056**	430	6.3	70236	58033	351180	20684	5.89

Table 6. An example of results from packet data simulations with CRRM (DR + IS-NCCRS, 6 layers, 320 kbit/s channel, 40 min simulation)

Note that DR rate is roughly 5/6 of total number of packet sessions, because the probability that other layer cell is less loaded than original cell is 5/6 (with 6 layers) in these simulations. This could be greatly reduced by introducing similar delay trigger and MinDelayDiff parameter as in case of IS-NCCRS. However, this could possibly reduce the CRRM gain down to same level as in 'IS-NCCRS only' case (see Figure 21 and Figure 22).

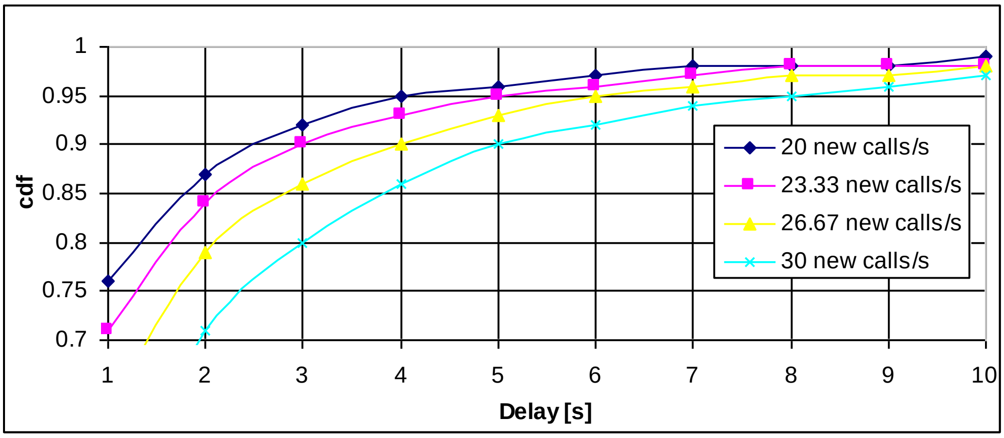


Figure 20. Example of cumulative distribution function of packet delay from simulations with CRRM utilising both IS-NCCRS and Initial Access cell reselection (6 Layers)

In following Figure 21 the system capacity and capacity improvements from different algorithms are plotted for interactive web surfing traffic through 320 kbit channel bandwidth, when 95% of users experience less or equal than 5seconds of delay.

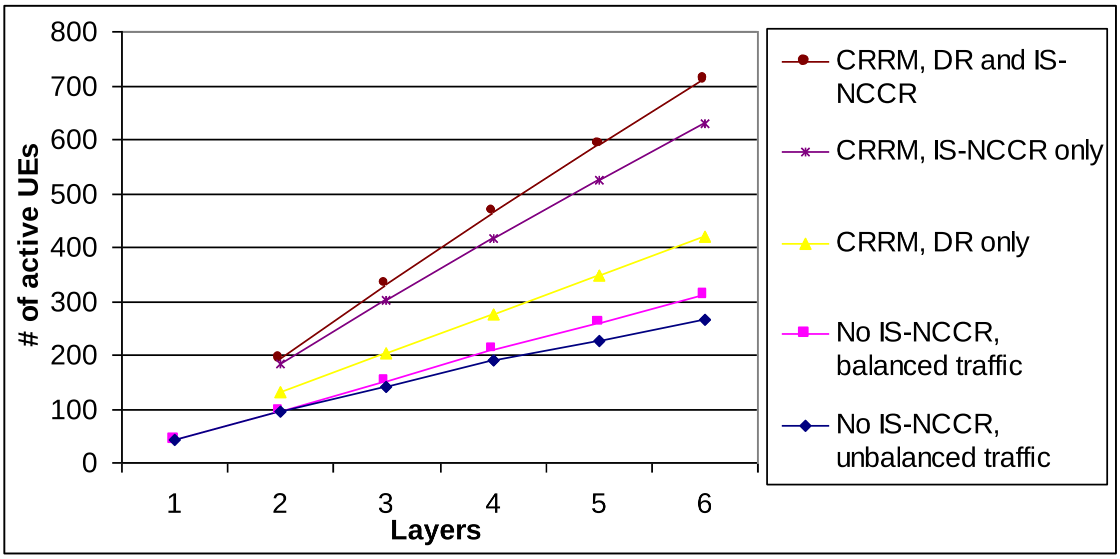


Figure 21. NRT simulations (channel bandwidth 320 kbps), 95% of users experience less than 5s delay

Also 90% outage values are plotted in Figure 22 in order to study how much the observed outage point affects the results.

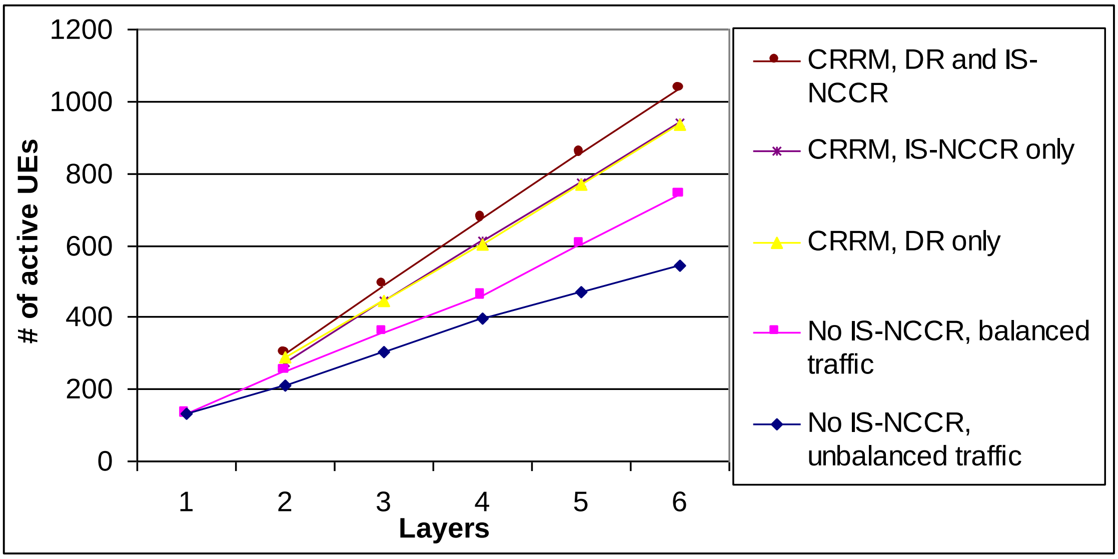


Figure 22. NRT simulations (channel bandwidth 320 kbps), 90% of users experience less than 5s delay

When comparing Figure 21 and Figure 22 it can be noted that the relative improvement by CRRM is higher with 95% outage than with 90%. The gain can be up to 150% with 95% outage and up to 100% with 90% outage. Thus, CRRM seems to provide more gain with stricter packet delay requirements.

6.3.1.2.3.4 Bandwidth 200kBps (1.6Mbps)

The same simulation were run with five times higher bandwidth (1.6Mbps). In Table 7 an example of results from CRRM simulations with both IS-NCCRS and DR (4 layers, 1.6 Mbit/s channel) is shown. Correspondingly, cumulative distribution function of packet delay for the same case is plotted in Figure 23. The number of active users at 95% outage (delay less or equal than 5s) is collected (**bold**) and plotted in Figure 24.

New calls/s	*Average actUEs*	*Users in buffer (ave)*	*Average buffer delay [s]*	*Number of packet sessions*	*Number of DRs*	*Number of packet calls*	*delay reason IS-NCCRS*	*IS-NCCRS rate %*
75	2043	499	1.3	40379	30540	201895	7288	3.61
91.67	2626	727	1.8	49399	37466	246995	13046	5.28
108.33	3384	1125	2.6	58607	44478	293035	21728	7.41
116.67	**3744**	1395	3.1	62207	47364	311035	26800	8.62
125	4801	2307	4.7	67613	51815	338065	39793	11.77

Table 7. An example of results from packet data simulations with CRRM (DR + IS-NCCRS, 4 layers, 1.6 Mbit/s channel, 10 min simulation)

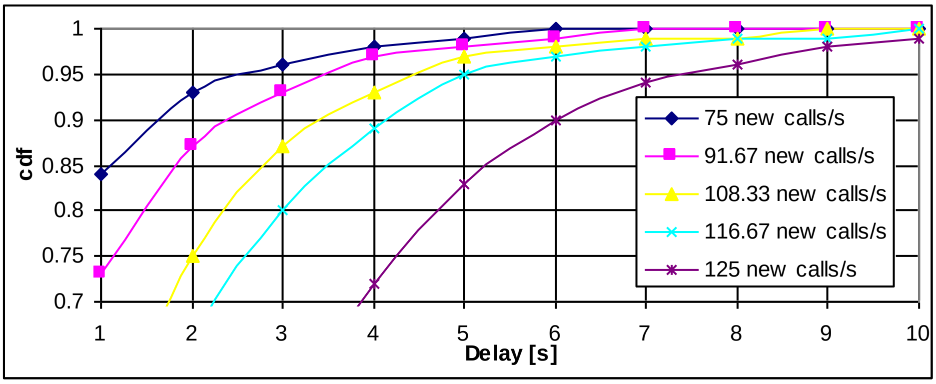


Figure 23. Example of cumulative distribution function of delay from simulations with CRRM utilising both IS-NCCRS and DR in cell reselection (4 Layers)

In Figure 24 the system capacity and capacity improvements from different algorithms are plotted for interactive web surfing traffic through 1.6 Mbps channel bandwidth.

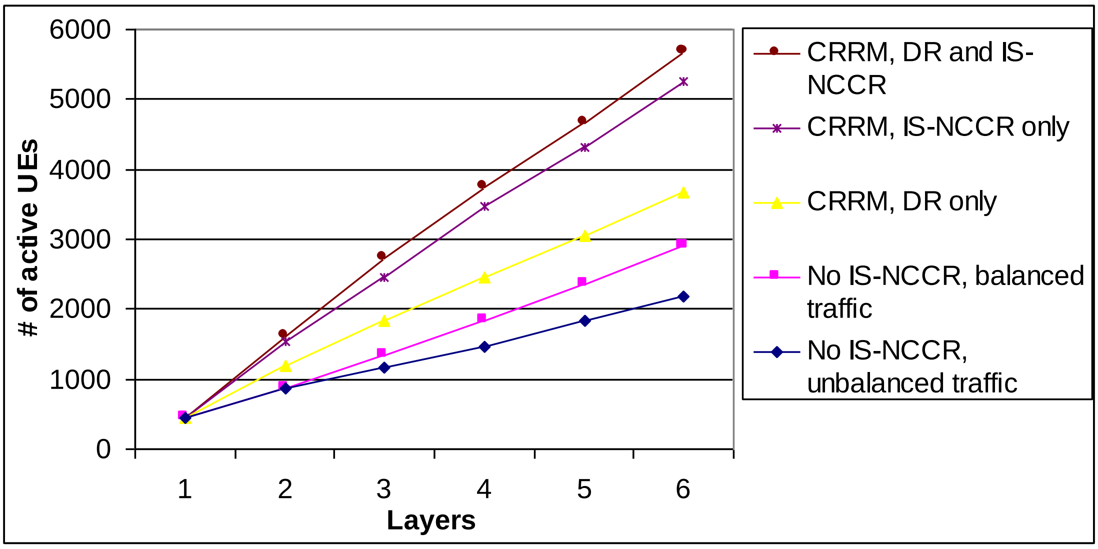


Figure 24. NRT simulations (channel BW 1.6 Mbps), 95% of users experience less than 5s delay

Again, when observing Figure 24 it is seen that the CRRM gain in 1.6Mbps channel can be up to 150% with 95% outage. Earlier in Figure 21 similar gains in 320 kbps channel were shown. Thus, it can be concluded that the throughput of one system does not heavily affect CRRM gain in case of NRT data.

6.3.1.2.3.5 Bandwidth efficiency with and without CRRM

In Table 8 an example of bandwidth efficiency of a cell with and without CRRM is shown for 4 layers. The CRRM efficiency (with both IS-NCCRS and DR) is compared with efficiency of no CRRM case.

The efficiency values are calculated by:

(NumOfNewUsers/s * AveKBpsPerUser * AveTxTimePerUser / NumberOfCells) / BandwidthPerCell

Channel Bandwidth	5s Outage Percentage	Bandwidth efficiency with CRRM	Bandwidth efficiency without CRRM
320 kbps	90%	0.61	0.34
	95%	0.47	0.18
1.6 Mbps	95%	0.73	0.31

Table 8. Bandwidth efficiency of a cell with and without CRRM, CRRM with IS-NCCRS + DR, no CRRM case with traffic unbalance, 4 layers

In these values the same CRRM capacity gain is seen as in Figure 21 - Figure 24. Also it is seen that 1.6 Mbps channel provide significant trunking gain as compared to 5 times smaller bandwidth.

6.3.1.2.4 Conclusions

The simulation results can be summarised in Table 9 as follows:

QoS class	Capacity gain with 2 layers	Capacity gain with 4 layers	Capacity gain with 6 layers
Conversational & Streaming	32 kbps: - 144 kbps: - 384 kbps: -	32 kbps: 4% 144 kbps: 11% 384 kbps: 27%	32 kbps: 6% 144 kbps: 16% 384 kbps: 40%
Interactive	40 - 100%	70 - 140%	90 – 180%

Table 9. Summary of CRRM gains for RT and NRT traffic classes

As a summary, the reason for CRRM gains are:

- RT traffic: only one DR is allowed for a user in call setup and handover, thus the gain comes from directing the call to most optimum cell/layer.
- NRT traffic: no load reason inter-system cell reselection used without CRRM

Note that CRRM gains for interactive non-real-time traffic depend heavily on required delay outage percentage, 90% vs. 95%.

CRRM gains for background traffic is more difficult to quantify. However, the average delay can be also minimised for background traffic, and thus the average throughput can be maximised, if the rest of the traffic is distributed equally between systems/layers.

It can be concluded that CRRM is most important for:

- high bit rate (>32 kbps) conversational and streaming connections
- interactive connections

It must be noted that the gains of these simulations are fairly ideal and assume no delays in signalling, etc.

Note 1: In the simulations, each layer has the same properties. As a result, the cost when moving between 2 layers is always identical. However in cases of multi-system networks (e.g. GSM/UTRAN), the cost (=signalling overhead, delay,..) for mobility in between layers of the same system (e.g. one UTRAN carrier to another carrier) is normally less significant than for mobility between layer of different systems.

Note 2. This simulation only considered the case of no CRRM for any layers or the case of CRRM between all layers. i.e. the case that CRRM is hierarchical is not considered. In case of an operator with a GSM and a UTRAN network, typically each system will already provide "CRRM" for the layers within that system.

6.3.1.2.5 IS-NCCR without CRRM(Additional Simulation)

In this section the comparison between IS-NCCR with and without CRRM is carried out. The basic differences between two concepts are:

- Blind selection of target IS-cell without CRRM. Source cell does not have any information of the target cell load. This may cause ping-pong effect between cells and high number of NCCRs, if the threshold is set to low value.

- Full control of resources with CRRM. NCCRs are only performed when seen feasible, target cell clearly less loaded than source cell. Thereby, the buffer load can be optimally distributed and less unnecessary NCCRs are performed. The number of buffer relocations and core network signalling is minimised.

In Figure 25 the comparison between concepts is shown in 320kbit shared channel. It is seen that the gain depends heavily on the parameters and the observation point chosen.

If the trigger of 5s delay threshold is used the capacity gain of IS-NCCR without CRRM is almost the same as with CRRM. This is due to the nature of the packet traffic model (heavy tailed Pareto distribution of packet call size) - it is rather probable that only one of the parallel layers is heavily delayed at a time. Therefore, as the threshold is set to rather low value (less than 5s) there is not a big difference if the target cell is selected randomly or optimally. However, the number of IS-NCCR is increased as later seen in Figure 26 and Figure 27.

A larger delay trigger (10s) was chosen to reduce the amount of IS-NCCRs. Those results are also seen in Figure 25. It can be noticed that CRRM controlled case provides 25% higher capacity than the case without CRRM control. However, the overall gain is less than in the case of 5s delay trigger.

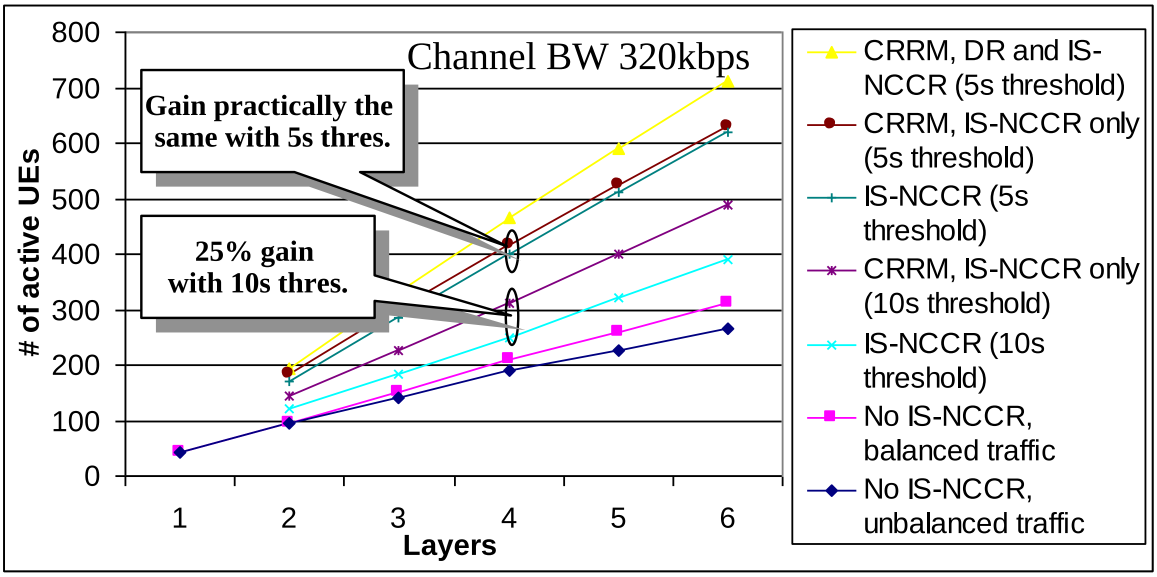


Figure 25. Comparison of IS-NCCR with and without CRRM (channel BW 320kbps), 95% of users experience less than 5s delay

Note. 5s delay already may not be acceptable for Internet service or user traffic.

In Figure 26 and Figure 27 the number of delay reason IS-NCCRs is presented with/without CRRM for two and four layers, respectively. It can be seen that the number of delay reason NCCRs is manifold without CRRM as compared to the CRRM controlled case due to the blind target decision.

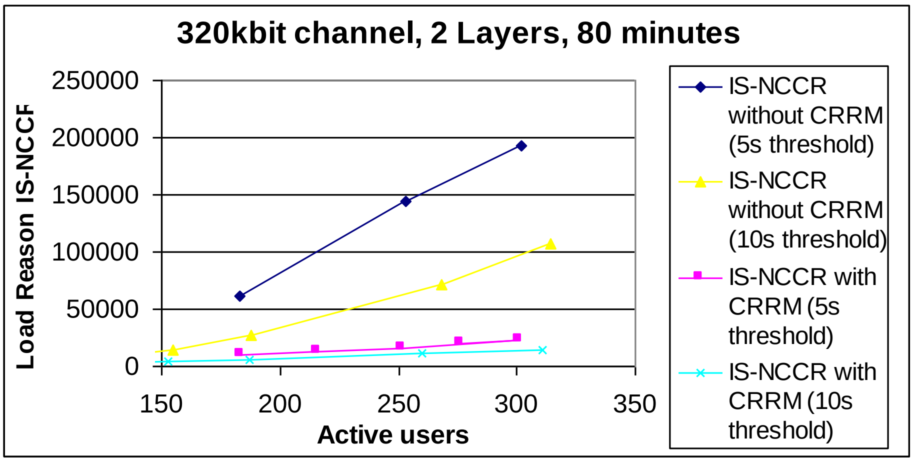


Figure 26. Comparison of the number of load reason IS-NCCRs with and without CRRM, 2 layers

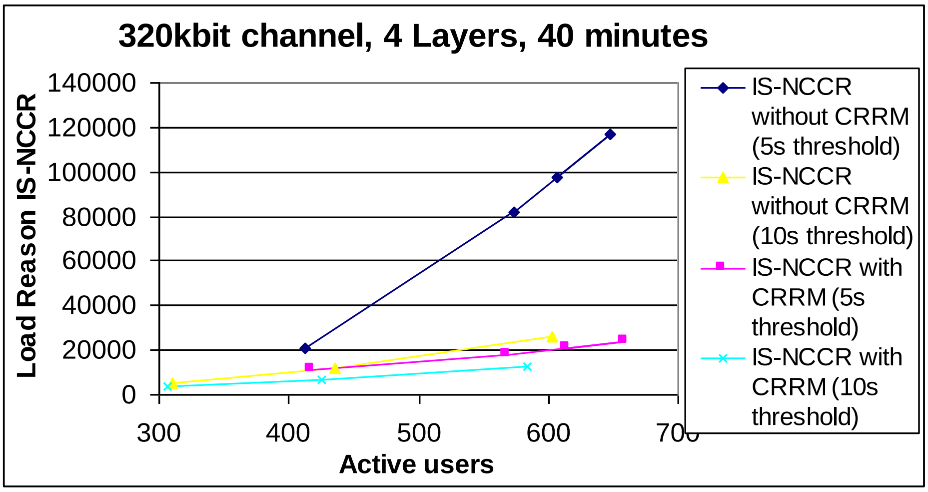


Figure 27. Comparison of the number of load reason IS-NCCRs with and without CRRM, 4 layers

As a **conclusion** it can be said that with low delay thresholds the capacity gain of IS-NCCR can be practically the same whether the CRRM is consulted or not. However, the use of low delay threshold without knowing the target cell causes significantly higher number of NCCRs than CRRM contro. This due to the blind decision vs. full control of resources.

If the delay threshold is increased to reduce the number of NCCRs, CRRM control results higher performance (25%) and still the number of NCCRs is significantly less with CRRM. However, the overall gain of IS-NCCR is reduced in both cases.

6.3.2 Simulation 2

6.3.2.1 Simulation model and assumptions

These simulations focus on TCP packet switched traffic, since most CRRM gains are found for PS traffic, according to chapter 6.3.1.2.4. This kind of traffic is dominating on Internet. Some congestion-avoiding mechanism, based on TCP, RTSP or other possible protocols, is essential for Internet.

For circuit-switched traffic the CRRM Air Interface capacity gain is virtually eliminated if the network is allowed to make several "Directed Retry" or "Handover" attempts, until a candidate cell with free capacity is found. If only a single "Directed Retry" is allowed, chapter 6.3.1.2.4 shows the expected gain. Since current GSM systems are usually configured to allow multiple "Directed Retry" attempts, the case with two layers, one for GSM and one for UMTS, is the most appropriate one. For packet switched traffic, however, it is more unclear what conditions are most realistic.

The simulation model is described in chapter 6.3.2.5.

The main differences compared to earlier simulations are:

- The traffic distribution mechanism is based on available (momentary) bandwidth per channel, not the channel buffer delay
- The peak available bandwidth is limited to 100kb/s

Some motives for this are:

- Bandwidth is probably a better measure than delay, since the subjectively tolerable delay relates to the volume of the downloaded data. If a user switches off images in web pages, he wants the delay to shrink in proportion to the size reduction.
- The length of each File (called Packet Call Size) is not always known. When TCP is used across the air interface, the performance will be severely impaired by large buffers. If there are only small buffers in UTRAN the knowledge of remaining length of an ongoing "Packet Call" is very small. A natural algorithm is then to distribute traffic based on "fair share" of instantaneous bandwidth
- TCP adapts the bit rate to the "channel" and the "saturation speed" after rate adaptation will be limited by MSS, RTT and Internet error rate. The rate 100kb/s may be pessimistic, but may compensate for the fact that the model is optimistic in other respects, e.g. neglecting the effects of non-instantaneous TCP rate adaptation to changes in the channel bandwidth. The limitation to a maximum bit rate thus accounts to some degree for the TCP behaviour.

6.3.2.2 QoS Measure

The method chosen to compare traffic performance is:

1. Select QoS measure(s) for the kinds of traffic involved and set a minimum QoS.
2. Simulate the system to find the maximum amount of traffic the system can handle while fulfilling the QoS limit.
3. Compare the amounts using different traffic algorithms

The outcome depends considerably on the QoS measure. In this report we use two QoS measures:

1. Average bandwidth > threshold
2. Probability that average bandwidth for one File Transfer is below threshold is smaller than P.

Both can be motivated for delay-insensitive traffic, where there is a compromise between air interface usage and user-perceived performance. The first does not consider the individual performance for a specific call, but only the average for all users. The second is more specific and accounts better for significant individual bandwidth variations, especially stalled channels. This may be a better measure of user satisfaction, since (human) users are more frustrated by significant deviations from expected bandwidths than if the bandwidth is 150 or 200kb/s on average.

6.3.2.3 Results

Figure 1 below shows the average user bit rate (system throughput) with six different algorithms at 80% system load, i.e. the offered load is 80% of the theoretical maximum capacity of $6 layers \times 9 cells \times 8 channels \times 100 kb/s = 43.2 Mb/s$

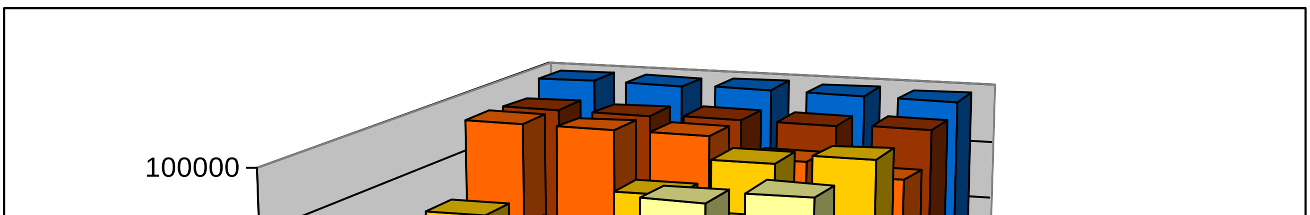
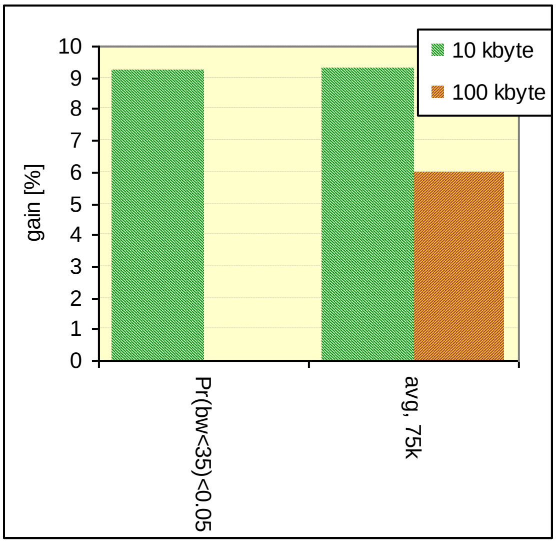


Figure 28. Average bitrate at 78% offered load

It can be seen that close to full user performance (set to 100kb/s) is achieved at this load.

Figure 29 shows the capacity gain at different minimum QoS measures.



Note: Bar #2 shows the value zero

Figure 29. Capacity gains of function [Bitrate, HHO] compared with [Balanced, noHHO] at 10and 100 kbyte file sizes, 6 layers.

The limits chosen may be justified as follows:

1 To achieve a good balance between user service and system utilisation, the average QoS should not be set too close to the maximum achievable (set to 100kb/s in this simulation). 75% is a reasonable value.

2 To achieve a good balance between user service and system utilisation, the threshold of minimum QoS should not be too close to the maximum achievable. The idea is to use low-QoS traffic as a "buffer" to utilize (almost) the full system capacity. The minimum QoS for this traffic should thus be quite low. We assume that 30-35kb/s is an upper bound on the lower threshold of QoS – maybe the threshold should even be below 5kb/s.

In general the capacity gain decreases with lower QoS limit. If there is a significant amount of best effort traffic accepting a low minimum rate, say 5kb/s, the gain would be significantly lower.

Note that the figures presented for 6 layers in this contribution are thus an upper bound on the gains to be expected.

Figure 30 and Figure 31 below, generated from the methods presented in W. Roberts "Realising QoS Guarantees in Multiservice Networks", show the principle:

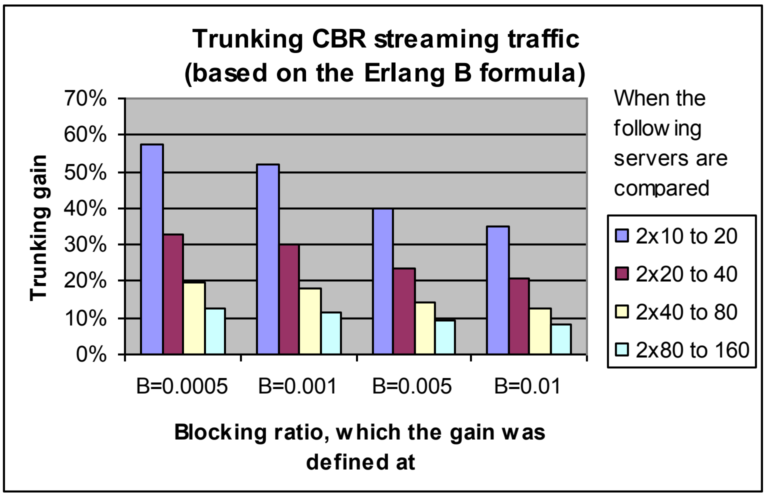


Figure 30. Aggregating streaming traffic from 2 identical servers

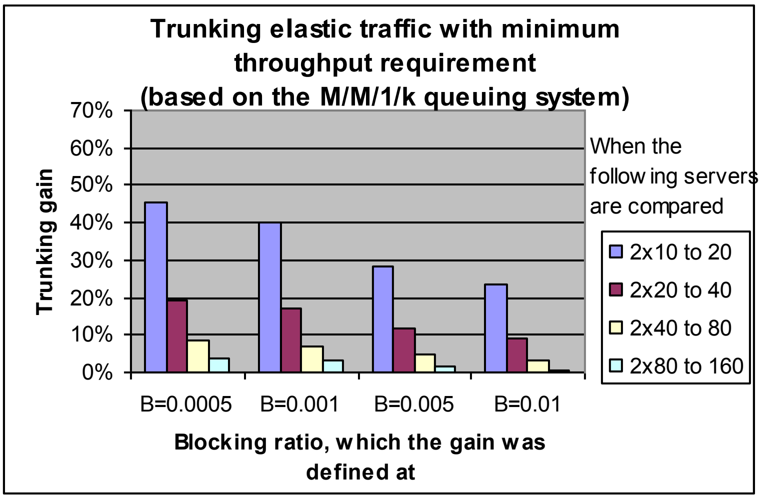


Figure 31. Aggregating elastic traffic from 2 identical servers

In practice GSM systems typically already implement some load sharing and UMTS systems will probably do so at least between cells of the same kind (e.g. co-located macrocells). So there will normally not be 6 "layers", but perhaps only two: GSM and UMTS. The capacity gains will then diminish. Figure 32 below shows the case for 2 layers with capacity 800kb/s in each.

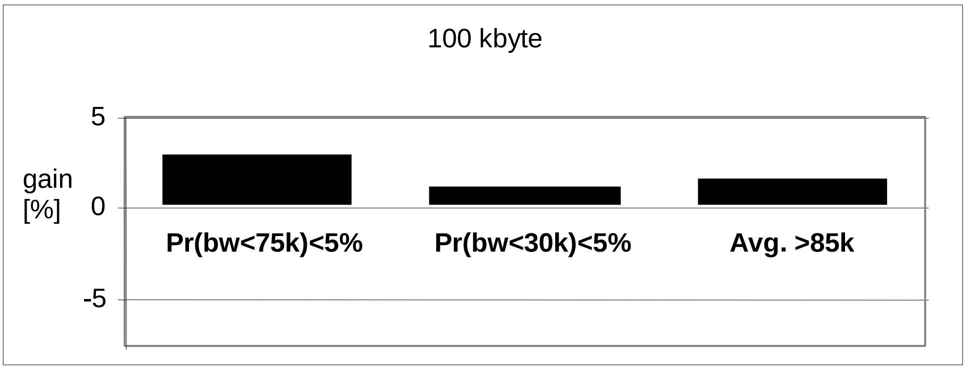


Figure 32. Capacity gains of function [Bitrate, HHO] compared with [Balanced, noHHO] at 10and 100 kbyte file sizes, 2 layers.

6.3.2.4 Conclusion

The capacity gain provided by CRRM for packet data is extremely dependent on the chosen QoS and traffic model.

If TCP is used and the performance limit is set to 30kb/s or lower, the CRRM gain will be <10%.

6.3.2.5 Description of the simulator

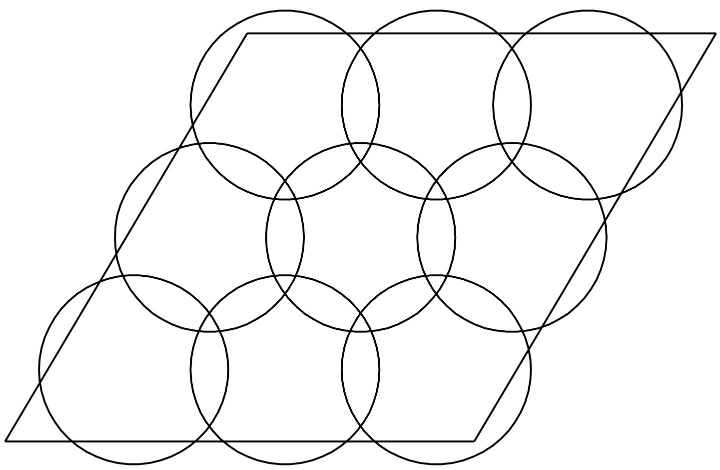
The simulator is simplified and does not include any "radio measurements". Cell (or "channel") selection for Directed Retry o r Handover is thus based on UE geographical position and cell loads.

6.3.2.5.1 Network topology

For simplicity all "layers" look the same. Although the layers are all called "UMTS" in the following, the results would be valid if some layers would be GSM/GPRS (assuming the capacity in the GPRS layer is the same). 2 or 6 layers are used. Six is a high number, but it should give the maximum CRRM gain.

The coverage per cell is assumed to be fixed – within a radius there is coverage, otherwise there is no coverage.

Figure 33. *Cells in a layer*



Cell parameters

System type Layer Frequency Cell Size	UMTS 0, 1, 2, 3, 4, 5 f0, f1, f2, f3, f4, f5 303 units
Base Station distance	181 units
Channels/cell (speech channels, not used in this simulation) Cell capacity Channels/internet call Bitrate/user (cell bitrate/8 channels)	64 800kb/s 8 100 kbit

The capacity figures are very approximate – exact values are not essential for the analysis.

The number of file downloads is greater than 10000 in all simulations. The simulated time varies from 300 to 10000 seconds depending on the file size.

6.3.2.5.2 Mobility model

Mobility parameters for all users	
turn angle per step	Uniform distribution [-10 deg, +10 deg]
speed	7 units/ sec
step time	1 sec
initial direction	uniform distribution [0 deg, 360 deg]
initial position	uniform distribution on the whole area

6.3.2.5.3 Traffic Model

This model shows the user behaviour. If the user perceives a bandwidth less than "minimum acceptable bandwidth" for "maximum waiting time", he will terminate the "call". In this simulation only one limit is used for all UE's, for simplicity.

Five different (fixed) file lengths are used, to test the dependency of "data call duration". All users transmit 200kbyte/hour in average.

"File length" can be a model of a web page or a part of a web page, if it consists of several objects.

Traffic model parameters for all calls	
service class Phone type	Internet UMTS/GSM
minimum acceptable bandwith	uniform distribution [15 kbps, 25 kbps]
maximum waiting time under acceptable bandwidth	uniform distribution [2 sec, 4 sec]

	case 1	Case 2	Case 3	case 4	case 5
call intensity [calls / hour] file length [byte]	0.02 10 M	0.2 1 M	2 100 K	20 10 K	200 1 K

6.3.2.5.4 Functionality

Three different algorithms are used at initial "data call setup", which is done for each file transfer:

1 "Unbalanced" means that "calls" are set up according to a predefined likelihood per "layer" This likelihood is not equal per layer, thus the word 'unbalanced'. Once the layer is chosen, the single cell closest to the UE is chosen. (This is not a desired case, but may illustrate an non-ideal case)

2 "Balanced" means that "calls" are set up according to a predefined likelihood per "layer". This likelihood is equal per layer, thus the word 'balanced'. Once the layer is chosen, the single cell closest to the UE is chosen

3 "Best bit rate" chooses a cell among those with coverage at the MS/UE, selecting the cell offering the best bit rate for this particular call. Since there is a fixed bandwidth of 800kb/s, the bandwidth per user becomes $800kb/s \big/ number\_of\_active\_users\_in\_cell$. If several cells offer the same bandwidth, a random one is picked among them.

Two different algorithms are used when the UE crosses the handover threshold while a file transfer is ongoing:

1 "No HHO". The UE stays in the layer originally chosen for the file transfer.

2 "HHO". The cell offering the best bit rate is chosen, i.e. a new layer can be chosen at any "handover"

It can be noted that Soft Handover is not implemented in UTRAN, for simplicity.

Traffic Management parameters			
initial cell selection strategy	Case 1	case 2	case 3
	Interlayer selection: unbalanced (L0:40.8%, L1:20.4%, L2:13.6%, L3:10.2%, L4:8.2%, L5:6.8% Intralayer selection: based on distance	Interlayer selection: random Intra-layer selection: based on distance	selection: best bitrate than random
cell selection at handover	Case 1	case 2	
	no interlayer handover	selection: best bitrate then random	

Some other potential, but not simulated, resource management functions not based on buffer delay, are:

1 UE selects the cell with the lowest load, based on CPICH Ec/No, before starting transmission. In this case the performance would probably be between "balanced" and "best bit rate", but closer to the first, due to rather slow response and less accurate measurements.

2 A handover can be initiated at “any time”, not only at "file transfer start" or "cell change", but any "severe load imbalance between layers". This method will probably give modest gains.

7 Agreements and associated contributions

8 Specification Impact and associated Change Requests

This section is intended to list the affected specifications and the related agreed Change Requests. It also lists the possible new specifications that may be needed for the completion of the Work Task.

9 Project Plan

9.1 Schedule

Date	Meeting	Scope	[expected] Input	[expected]Output

9.2 Work Task Status

	Planned Date	Milestone	Status
1.			
2.			

History

Document history		
Date	**Version**	**Comment**
07/2001	0.1.0	TR template
08/2001	0.2.0	Introduction, requirement, approach 1 are added
08/2001	0.3.0	Approach 2 and Simulation result are added
10/2001	0.4.0	Detail information of approach 1 was added.
11/2001	0.4.1	2 more additional simulations were added. TR was restructured.-Common measurement part was moved as common part of all solutions. One more solution was introduced. Chapter about 2nd approach was restructure. Clarification was made on TR.
12.2001	2.0.0	This has been made version 2.0.0.
Editor for 3G TR 25.881 is Woonhee Hwang, Nokia :		
Tel. : +358 9 5113 0149 Fax : +358 9 5113 0163 Email : woonhee.hwang@nokia.com		
This document is written in Microsoft Word version.		

Annex A (informative): Change history

Change history							
Date	TSG #	TSG Doc.	CR	Rev	Subject/Comment	Old	New
12/2001	14	RP-010923	-	-	Approved at TSG RAN #14 and placed under Change Control	-	5.0.0