

Grant Agreement No.: 671705 Research and Innovation action Call Identifier: H2020-ICT-2014-2





quality of Service Provision and capacity Expansion through Extended-DSA for 5G

D4.2: RM framework and modelling

Deliverable type	Report	
Dissemination level	PU	
Due date	2017-03-31	
Submission date	2017-04-19	
Lead editor	Shahid Mumtaz (IT)	
Authors	Keith Briggs, Salvador Diaz (BT); Antonio de Domenico (CEA); Shahid Mumtaz, Kazi Saidul Huq, Imran Khan, Paulo Marques, Jonathan Rodriquez (IT); Faouzi Bouali, Dionysia Triantafyllopoulou, Haeyoung Lee (UNIS); Ioannis-Prodromos Belikaidis, Voula Vassaki, Andreas Georgakopoulos, Evangelos Kosmatos, Evangelia Tzifa, Kostas Tsagkaris, Aikaterini Demesticha, Vera Stavroulaki, Panagiotis Vlacheas, Dimitrios Kelaidonis, Yiouli Kritikou, Athena Ropodi, Panagiotis Demestichas, Marinos Galiatsatos, Aristotelis Margaris, George Paitaris, Dimitrios Kardaris, Evangelos Argoudelis (WINGS).	
Reviewers	Valerio Frascolla (Intel), Michael Fitch (BT).	
Work package, Task	WP4, T4.2	
Keywords	System-level simulation, interfaces and entities, RRM demulator, small cells	

Abstract

This deliverable describes the radio resource management (RRM) framework and modelling for integrating and assessing the proposed algorithms in SPEED5G eDSA framework. The focus is on RRM and interfaces to the MAC layer and also northbound interfaces to the OAM, spectrum manager, and KPI manager. This deliverable also contains the modelling of demulator for initial testing for algorithms before it is implemented in a full hardware testbed.



Version	Date	Description of change	List of contributor(s)
v0.1	2017-01-26	Template/ToC draft	Shahid Mumtaz
v0.2	2017-02-01	Updates to template	Seiamak Vahid
v0.3	2017-02-08	RRM chapter	Salva Diaz, Keith Briggs
v0.4	2017-02-27	RRM revisions	Salva Diaz, Keith Briggs
v0.5-0.8	2017-03-03	Chapter rearrangements and revisions; new figures	Keith Briggs
v0.9	2017-03-08	Chapter 2 updates	ВТ
v0.10	2017-03-10	Chapter 2.6 and Chapter 4 updates	WINGS
v0.11	2017-03-13	Editing of chapter 4	Keith Briggs
v0.12	2017-03-13	Chapter 4 updates	WINGS
v0.13	2017-03-13		
v0.14	2017-03-14	Chapter 4 updates	WINGS
v0.15	2017-03-14	Chapter 2 updates	Keith, using comments from Mike
v0.16	2017-03-16	Algorithms received from involved partners	Shahid
v0.17	2017-03-16	revised	Shahid
v1.0	2017-03-17	English checks and revisions for consistency	Keith

Document revision history

Disclaimer

This report contains material which is the copyright of certain SPEED-5G Consortium Parties and may not be reproduced or copied without permission.

All SPEED-5G Consortium Parties have agreed to publication of this report, the content of which is licensed under a Creative Commons Attribution-Non-commercial-NoDerivs 3.0 Unported License¹.

Copyright notice

© 2015 - 2017 SPEED-5G Consortium Parties

¹ http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en_US



Executive Summary

The RRM (Radio Resource Management) entity is a key component of the SPEED-5G architecture. It consists of two parts, the cRRM (centralised part) and dRRM (distributed part). A network operator will have one or more cRRM entities located typically located in a gateway entity of the network controlling several hundred cells, whereas the dRRM is located in every cell and will communicate with one or more cRRMs. The dRRMs can communicate with cRRMs from different operators and in this way the cell becomes multi-tenanted.

This deliverable defines the RRM framework including the cRRM with its interfaces to other parts of the core network (spectrum manager, KPI collector and OSS) and to dRRMs, and the dRRM with its interfaces to the cRRM and to the MAC layer The proposed RRM design is capable of incorporating algorithms from multiple vendors whether those algorithms are centralized or distributed. Moreover, the proposed RRM framework fully decouples the underlying algorithms by the introduction of an abstraction layer (AL) and supports multiple interfaces transparently from the algorithmic point of view. This means that communications with the entities outside the RRM are the responsibility of the AL. The framework can also work in Cloud Computing environments, as well as in embedded equipment.

This deliverable also incorporates several RRM algorithms from the SPEED-5G consortium, which will eventually be integrated into the proposed RRM framework. More specifically:

Algorithm 1: Efficient licensed-assisted access operation in same call cell based on reinforcement learning:

This algorithm is designed for operation in dense heterogeneous cellular networks with Licensed-Assisted Access (LAA) small cells, capable of operating in both licensed and unlicensed spectrum.

Algorithm 2: This algorithm is related to the RAT/spectrum/channel selection based on machine learning

This algorithm comprises RAT, spectrum and channel selection based on machine learning and takes into account the 3.5 GHz band for achieving better performance, especially in dense and congested 5G environments.

Algorithm 3: Radio resource allocation with aggregation for mixed traffic in a WiFi coexisted heterogeneous network

This algorithm is intended for resource allocation with aggregation to support different level of quality of service (QoS) of different traffic types in the cellular network where the WiFi network coexists.

Algorithm 4: Fuzzy MADM strategy for spectrum management in multi-RAT environments

This algorithm provides context-aware offloading in dense small cell environments to support a mixture of delay-sensitive and best-effort applications.

Algorithm 5: Co-primary spectrum sharing in uplink SC-FDMA networks

This algorithm allows mobile network operators (MNOs) employing infrastructure sharing to efficiently share the available spectrum resources, taking advantage of information coming from the Physical and MAC layers, in order to avoid inter-operator interference and achieve improved Quality of Service (QoS) for real-time applications

Algorithm 6: Dynamic resource allocation algorithms for coexistence of LTE-U and WiFi

This algorithm provides resource allocation for co-existing LTE-U and WiFi networks to maximize throughput and hence minimize the interference.

Lastly, this deliverable describes a demulator (demonstrator+emulator) software design. The proposed design will be tested in software before it is implemented in a full hardware testbed. This is



why the concept of a demulator is proposed. The demulator thus becomes in effect a reference model for the RRM framework, hosting different RRM-related algorithms enabling eDSA.

Table of Contents

Executive Summary			
Table of	Contents	. 5	
List of Fig	gures	. 8	
List of Ta	bles	10	
Abbrevia	tions	11	
1	SPEED-5G Radio Resource Management Framework	13	
1.1	Design goals	13	
1.2	Framework design	13	
1.3	Abstraction layer	16	
1.3.1	RRM Configuration	16	
1.3.2	Algorithms	17	
1.3.3	Requester	19	
1.3.4	Internal KPI collector	19	
1.3.5	Message constructor	20	
1.4	Interfaces	22	
1.5	Interfaces with the centralized RRM	22	
1.5.1	5G RRC	23	
1.5.2	MAC	23	
1.5.3	KPI Collector	23	
1.5.4	Spectrum Manager	24	
1.5.5	OAM/OSS	24	
1.5.6	Distributed RRM	24	
1.5.7	Centralized RRM	25	
1.6	Overview of RRM algorithms and functions	25	
1.7	Demulator	26	
1.7.1	Software Design aspects	27	
1.8	Tables defining messages below the abstraction layer	27	
2	RRM algorithm 1: Efficient licensed-assisted access (LAA) operation in small cells, based on reinforcement learning	30	
2.1	Relation to Golden Nugget	30	
2.2	Assumptions and system model	31	
2.3	Algorithm Description	31	
2.3.1	Algorithm pseudo-code	32	
2.3.2	Inputs and Outputs	32	
2.3.3	Simulation assumptions and parameters	34	
2.3.4	Performance evaluation	34	



3	RRM algorithm 2: RAT/spectrum/channel selection based on hierarchical machine learning	36
3.1	Relation to Golden Nugget	36
3.2	Assumptions and system model	37
3.3	Algorithm Description	38
3.3.1	Algorithm flowchart	38
3.3.2	Inputs & Outputs	40
3.3.3	Performance measures and KPIs	42
3.3.4	Simulation assumptions and parameters	42
3.3.5	Performance Evaluation	43
4	RRM algorithm 3: Radio resource allocation with aggregation for mixed traffic in a W coexisted heterogeneous network	iFi 45
4.1	Relation to Golden Nugget	45
4.2	Assumptions and system model	46
4.3	Algorithm description	47
4.3.1	Algorithm flowchart	47
4.3.2	Inputs and Outputs	49
4.3.3	Simulation assumptions and parameters	51
4.3.4	Performance Evaluation	52
5	RRM algorithm 4: Fuzzy MADM strategy for spectrum management in multi-RAT environments	54
5.1	Relation to Golden Nuggets	54
5.2	Assumptions and system model	55
5.3	Proposed context-aware user-driven framework	56
5.3.1	Functional architecture	56
5.3.2	Integration into the SPEED-5G architecture	57
5.4	Connection manager: Fuzzy MADM decision-making	58
5.4.1	Preliminary target behaviour	59
5.4.2	Out-of-context suitability levels	59
5.4.3	In-context suitability levels	60
5.4.4	Decision-making	62
5.4.5	Inputs and Outputs	62
5.4.6	Traffic mixture and performance measures/KPIs	65
5.4.7	Simulation assumptions and parameters	65
5.4.8	Benchmarking	65
5.4.9	Performance evaluation	66
6	RRM algorithm 5: Co-primary spectrum sharing in uplink SC-FDMA networks	68
61	Relation to Golden Nugget	68

Speed 55

6.2	Assumptions and system model	68
6.3	Algorithm Description	68
6.3.1	Algorithm flowchart	68
6.3.2	Inputs & Outputs	72
6.3.3	Performance measures/KPIs	74
6.3.4	Simulation assumptions and parameters	74
6.3.5	Performance Evaluation	75
7	RRM algorithm 6: Dynamic resource allocation algorithms for coexistence of LTE-U and WiFi	79
7.1	Relation to GN	79
7.2	Assumptions and system model	80
7.2.1	System Model	80
7.2.2	Assumption	80
7.2.3	Achievable data rate in unlicensed spectrum	82
7.2.4	Achievable data rate in licensed spectrum	83
7.2.5	Problem formulation	83
7.2.6	Algorithm description	84
7.2.7	Algorithm pseudo-code	85
7.2.8	Algorithm flowchart	86
7.2.9	Inputs and outputs	86
7.2.10	Performance measures/KPIs	88
7.2.11	Simulation assumptions and parameters	88
7.2.12	Performance evaluation	90
8	Conclusions	92
Referenc	zes	94
Appendix	x A Fronthaul/backhaul requirements for small cells when deployed on the SPEED-5G use-cases	96
A.1	Location of the fronthaul	96
A.1.1	Below the PHY layer	97
A.1.2	At the MAC layer	97
A.1.3	At the PDCP layer	98
A.2	Practical measurements	99
A.2.1	Test 1 – Direct Ethernet connection	99
A.2.2	Test 2 – With G.FAST	00



List of Figures

Figure 1: RRM overview
Figure 2: High-level RRM framework. Note the abstraction layer, a critical component of the design separating higher-level algorithms from internal components. The colour-coding for the entities is maintained on subsequent figures
Figure 3: SPEED-5G RRM algorithms 17
Figure 4: Algorithm structure
Figure 5: SW abstraction of the Message constructor
Figure 6: SPEED-5G 5G Greenfield network architecture [1] 22
Figure 7: SPEED-5G Protocol Stack architecture [1]
Figure 8: The heterogeneous network under investigation
Figure 9: Mapping of the efficient LAA operation in small cells to CRRM/MAC blocks
Figure 10: Message sequence chart for the efficient LAA operation in small cells, based on reinforcement learning
Figure 11: Probability of successful access in the LAA band with the Multi-armed Bandit and a random access scheme
Figure 12: Cumulative Distribution Function of the user latency with the Multi-armed Bandit and a random access scheme
Figure 13: Cumulative distribution function of the user throughput with the Multi-armed Bandit and a random access scheme
Figure 14: Centralized vs. distributed management
Figure 15: SAS three-tier Model
Figure 16: 3.5 GHz spectrum allocation for each tier of the SAS model
Figure 17: Flowchart of algorithm with learning capabilities as an option
Figure 18: Mapping to RRM functional blocks as defined in SPEED-5G D4.1
Figure 19: Message sequence chart for RAT/spectrum/channel selection based on hierarchical machine-learning
Figure 20: a) Average session duration, b) Average cell-edge throughput (5%)
Figure 21: An overview of LTE-WiFi Aggregation (LWA) [14]45
Figure 22: Support for multiple users of heterogeneous traffic types with the use of carriers in multiple licensed bands and WiFi network
Figure 23: Curves of utility functions with different parameters for heterogeneous traffic (blue colour for inelastic traffic and purples colour for elastic traffic)
Figure 24: The flowchart of the proposed resource allocation algorithm
Figure 25: Mapping the proposed algorithm to RRM functional blocks of SPEED-5G
Figure 26: Message sequence chart of the proposed resource allocation algorithm
Figure 27: Average utility performance for different number of UEs
Figure 28: Average data rate achievable from WiFi for different UE speeds
Figure 29: Illustrative example of SINR map

Speed55

Figure 30: Functional architecture of the proposed context-aware user-driven framework
Figure 31: Integration into the SPEED-5G architecture
Figure 32: Flowchart of the proposed fuzzy MADM strategy
Figure 33: WLAN/VoIP fuzzy logic controller 60
Figure 34: LTE/VoIP fuzzy logic controller
Figure 35: Message sequence chart of the relevant exchanges for the fuzzy MADM strategy
Figure 36: Evolution of the end-to-end delay of VoIP
Figure 37: Analysis of the effectiveness in achieving the target behaviour
Figure 38: Flowchart of the proposed uplink co-primary spectrum sharing algorithm in each Time Transmission Interval (TTI)
Figure 39: Flowchart of the calculation of <i>Gi</i> , <i>k</i>
Figure 40: Message Sequence Chart of the uplink co-primary spectrum sharing algorithm
Figure 41: Average packet timeout rate of (a) MNO1 and (b) MNO2 versus the number of users and the shared spectrum access priority
Figure 42: Average delay of (a) MNO1 and (b) MNO2 versus the number of users and the shared spectrum access priority
Figure 43: Goodput of (a) MNO1 and (b) MNO2 versus the number of users and the shared spectrum access priority
Figure 44: Fairness (Jain index) versus the total number of users and the MNO priority
Figure 45: Dynamic resource allocation algorithms scenario
Figure 46: Algorithm flowchart
Figure 47: Mapping to RRM functional blocks as defined in SPEED-5G D4.1.
Figure 48: Message sequence chart
Figure 49: (a) Downlink user data rate of each network over total served traffic per AP, (b) 5th percentile SINR of operator A with two energy-detection thresholds
Figure 50: Ideal and non-ideal fronthaul
Figure 51: Definition of fronthaul and backhaul
Figure 52: SPEED-5G small cell stack with fronthaul below PHY layer
Figure 53: Fronthaul between MAC and PHY
Figure 54: Fronthaul between PDCP and RLC
Figure 55: Test equipment layout and latency results for direct Ethernet connection
Figure 56: Extension of test using 100m of copper pair running G.FAST



List of Tables

Table 1: Summary of algorithms	. 26
Table 2: Summary of RRM interfaces defined in D4.1	. 29
Table 3: Interface options between dRRM and cRRM	. 29
Table 4: List of INPUT parameters used by the efficient LAA operation in small cells.	. 32
Table 5: List of OUTPUT parameters/values provided by the efficient LAA operation in small cells	. 33
Table 6: Simulation assumptions and parameters of proposed algorithm	. 34
Table 7: Input parameters of proposed algorithm	. 40
Table 8: Output parameters of proposed algorithm	. 40
Table 9: Performance measurements and KPIs of proposed algorithm	. 42
Table 10: Simulation assumptions and parameters of proposed algorithm	. 43
Table 11: General simulation assumptions and parameters	. 43
Table 12: Tested scenario cases	. 44
Table 13: Input parameters of the proposed algorithm	. 50
Table 14: Output parameters of the proposed algorithm	. 50
Table 15: List of simulation parameters for the proposed algorithm	. 52
Table 16: Input parameters of the proposed fuzzy MADM strategy	. 63
Table 17: Output parameters of the proposed fuzzy MADM strategy	. 63
Table 18: Simulation assumptions and parameters for evaluating the fuzzy MADM strategy	. 65
Table 19: Uplink so-primary spectrum sharing algorithm pseudo code	. 72
Table 20: Input parameters of the uplink co-primary resource allocation algorithm	. 73
Table 21 Output parameters of the uplink co-primary resource allocation algorithm	. 73
Table 22: Simulation Parameters of the Uplink Co-Primary Spectrum Sharing Algorithm	. 75
Table 23: Notations	. 81
Table 24: Input parameters of proposed algorithm	. 86
Table 25: Output parameters of proposed algorithm	. 87
Table 26: Performance measurements and KPIs of proposed algorithm	. 88
Table 27: Simulation parameter LTE-U	. 89
Table 28: Operator WiFi Parameter	. 89
Table 29: Latency assumptions on fronthaul and backhaul	. 96
Table 30: Summary of fronthaul location pros and cons	. 99



Abbreviations

AL	Abstraction Layer
СА	Carrier Aggregation
СМ	Connection Manager
cRRM	Centralized Radio Resource Management
DCF	Distributed Coordination Function
dRRM	Distributed Radio Resource Management
FBMC	Filter Bank Multicarrier
FLC	Fuzzy Logic Controller
GAA	General Authorized Access
GPS	Generalized Pattern Search
GW	GateWay
hMAC	Higher MAC
НО	Hand-Over
HW	Hardware
IPSec	Internet Protocol Security
ISM	Industrial, Scientific and Medical
КРІ	Key Performance Indicator
КРІ	Key Performance Indicator
LAA	Licensed-Assisted Access
LLS	Link Level Simulation
LSA	Licensed Spectrum Access
LTE-U	LTE in Unlicensed band
LWA	LTE-WiFi Aggregation
MAC	Medium Access Control
MADM	Multiple Attribute Decision Making
MADS	Mesh Adaptive Direct Search Algorithm
MNO	Mobile Network Operator
MP	Multiprocessing
ΟΑΜ	Operation And Management
OMA-DM	Open Mobile Alliance-Device Management
OSS	Operations Support System
PAL	Priority Access Licenses
Phy	Physical interface
QoS	Quality of Service



QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RAT	Radio Access Terminal
RLC	Radio Link Controller
RPC	Remote Procedure Call
RRC	Radio Resource Controller
RRM	Radio Resource Management
SAS	Spectrum Access System
SCTP	Stream Control Transmission Protocol
SLS	System Level Simulator
SM	Spectrum Mobility
SOM	Self-Organizing Maps
SON	Self-Organized Network
SS	Spectrum Selection
SW	Software
UCB	Upper Confidence Bound
UE	User Equipment
USRP	Universal Software Radio Peripheral
Video over LTE	VILTE
Voice over LTE	Volte



1 SPEED-5G Radio Resource Management Framework

In this section, SPEED-5G RRM framework functionalities are defined including each of their entities. Then, the interfaces between RRM and the other components are specified, as well as the messages sent through each of the mentioned interfaces.

1.1 Design goals

The design of the SPEED-5G RRM framework must be capable of supporting diverse algorithms from multiple vendors irrespective of whether the solution is centralized or distributed. In order to do this, the proposed framework and the algorithms are completely decoupled. The decoupling is achieved by automatic mechanisms for algorithm addition, removal or update operations supported by the framework. More details about built-in automated mechanisms in support of aforementioned operations can be found in section 1.3.2.

The framework and algorithms decoupling allows the same framework to be used in a centralized or distributed environment. Depending on where it is deployed (remote or local), HW and SW restrictions may differ.

The framework design allows asynchronous mechanisms so that algorithms may run in parallel. Depending on the algorithm requirements, parallelisation may be done during the whole decision process; but the framework is also capable of running the algorithms in sequential manner when the outputs of some are used as inputs to others. This particular framework structure has been design for support of the enhanced RRM algorithms proposed in SPEED-5G.

The main characteristics of the proposed RRM framework design are:

- The framework is fully decoupled from algorithms, via an abstraction layer (AL).
- There is a single framework for both centralized and distributed versions of the algorithms.
- Multiple interfaces are transparently supported, from the algorithms point of view.
- A container for any data provided, or required by, the algorithms is provided.
- System reliability is ensured.
- Virtualization can be supported.
- Asynchronous procedures are allowed.
- The system is fully configurable via OAM/OSS.

1.2 Framework design

The most important design requirements built-into the framework, are:

- Cells only have one cRRM per operator.
- Cells support RAN-sharing.
- dRRMs may be connected with more than one cRRM, one per operator.
- A new interface between cRRMs is provided, to support RAN-Sharing.
- The only cRRM is capable of requesting data from the Spectrum Manager. Any dRMM request goes through the cRRM.
- RRM is composed of one cRRM and one or more dRRMs.
- Decisions taken at cRRM have priority over those taken at dRRM. When the cRRM sends a new configuration, the dRRM resets its status with cRRM information.



- dRRM is always deployed in a cell, even if it is only a gateway between cRRM and cell layers.
- cRRM is always deployed in a network, even if it is only a gateway between the Spectrum Manager and the cell.
- SON and current 3GPP RRM procedures are part of the RRM algorithms.
- Ready for Cloud Computing, through its flexible degree of distributed processing.



Figure 1: RRM overview

Figure 1 shows the proposed RRM architecture where a dRRM and a cRRM is deployed.

The high-level design of the RRM framework is shown in Figure 2, where the main blocks are:

- RRM configuration: storage of the full RRM configuration, including, e.g., framework procedures, algorithms order, and IP direction.
- Requester: used by the algorithms to get data periodically or on-demand. It does so by scheduling its periodic request, or by issuing requests when the algorithm runs.
- Internal KPI collector: storage of own RRM KPIs.
- Message constructor: gets the internal message and fills the required algorithms' structures.
- Abstraction layer: manages the physical interfaces and translates physical messages through its SW abstraction layer.
- Interfaces: responsible for supporting the different message protocols. See SPEED-5G architecture [1].
- Algorithms: In addition to supporting legacy RRM algorithms, the new algorithms proposed by SPEED-5G for an optimal spectrum usage. All algorithms are intended to be compatible with existing SON algorithms (which then run independently and in parallel), but the existence of SON is not assumed.

The full configuration setup of RRM is stored into the *RRM configuration* entity. This entity contains various parameters; for example, required IP addresses, algorithm order (when executed in series), different groups of algorithms to be executed in series, priority of the algorithm when more than one targets the same issue, etc. The *RRM configuration* is managed by the OAM, which in turn is controlled by the network operator.





Figure 2: High-level RRM framework. Note the abstraction layer, a critical component of the design separating higher-level algorithms from internal components. The colour-coding for the entities is maintained on subsequent figures.

The *Algorithms* entity contains all the RRM algorithms, including the Self-Organizing Network (SON) algorithms required in heterogeneous network deployments, and the algorithms defined in SPEED-5G to enhance the spectrum usage. The algorithms may be grouped and the framework has the capability to run each defined group in parallel. The groups of algorithms or the algorithms themselves are isolated and, as a consequence, each one will provide the best solution. Where more than one group provides a solution for the same case, Artificial Intelligence (AI) techniques can be deployed to make a final selection, so to further improve the system performance.

The *Message constructor* entity is responsible for validating the received data used by the algorithms and for forwarding a message to the message target. The validation process is done by the algorithm itself in what we call Data Inspector and Filters. This is necessary if decoupling between the framework and the algorithms is to be maintained. More details may be found in section 1.3.2. This entity is a key element in the RRM framework, as it allows speeding-up of the final process decision.

The *Interfaces* entity is in charge of managing the physical interfaces required by algorithms allocated in the RRM. Below this entity, the *Abstraction layer* entity provides the required isolation between the standardized messages and the SW abstracted messages.

The *KPIs collector* entity stores all the RRM indicators required by the operator. In a similar way to other entities in the system, the cRRM sends recieved KPIs to the Internal *KPI Collector* entity. The RRM KPI message periodicity for each KPI is defined by the OAM.



Finally, the *Requester* entity supports two different functionalities. On-demand services are used by the algorithm when more data is required after receive the initial trigger. On the other hand, the periodic requester is an optional entity, depending on the implementation of the external components. For example, if the *KPI Collector* supports periodic message scheduling, this entity is not necessary for that purpose.

1.3 Abstraction layer

The key idea here is to separate out the higher functions (especially cRRM, but also dRRM algorithmic operations), from the lower layers (RLC, hMAC, etc.) with an Abstraction Layer (AL). The AL translates all messages passing through it from a SW abstraction viewpoint, to HW-specific message primitives for the layers below it. The main design goal is to achieve flexibility in the architecture, in order to support with one system all the testbed configurations which are proposed in D6.1. The advantages are:

- Algorithm developers can design and develop RRM algorithms in purely high-level terms. Only the programmer of the AL is responsible for coding the translation into primitives. Also, the mapping onto hardware IP addresses and port numbers would be handled by the AL, and obviously this would be instantiated differently for the different testbed configurations.
- Only one RRM architecture is needed above the AL.
- Below the AL, different HW such as USRP, FBMC, etc. can be supported, with differences hidden from the higher layers.
- The main software concepts in the current RRM demulator; namely asynchronous processes for each RRM entity, and abstract buffered duplex queues for messaging between entities. This means that all RRM algorithms have to be constructed in such a way that they wait for inputs, perform computations, provide outputs, and then go back to waiting state for more inputs. Communication with the RRM is done via remote procedure call (RPC) protocol, which makes it easy to host the RRM remotely.
- If some algorithm needs to run closer to the HW than this architecture would allow, bypass operations are supported, as special function calls which can skip one or more layers.
- It has to be emphasized that SPEED-5G RRM design can also be implemented in a full network function virtualization environment (like SONATA http://sonata-nfv.eu/), and although such functionality will not be demonstrated, nevertheless the proposed RRM framework architecture is compatible and able to operate in virtualized environments.

1.3.1 RRM Configuration

This module contains the entire RRM configuration required for the functioning of the system. The configuration module contains all the information related to the configuration, for instance, how the temporal data is stored, in a database or RAM memory; the allowed protocol interfaces, or the authorized algorithms. The information required for secure connection to the network is also stored in this module.

Once the RRM is connected into the network, it is able to communicate with many other entities. For this reason, the IP address of each entity has to be provided by the OAM during the bootstrap. In a real deployment, this will need an IPSec tunnel establishment, so there will be an authentication and key exchange process. Once the RRM is running, the network operator is able to modify the default RRM configuration. The RRM interface to the OAM is via the TR-69 interface. Finally, due to the algorithm requirements, the RRM entity is connected to the KPI Collector, with the network cells, the Spectrum Manager, and other RRMs.

The proposed framework allows for the grouping of different algorithms. Several groups of algorithms may be created simultaneously and executed in parallel. In each group, the algorithms are executed in a defined sequence. The sequence can be configured by the operator through the OAM.

1.3.2 Algorithms

This section focusses on the SPEED-5G algorithm support provided by the proposed framework. Current state-of-the-art and SON algorithms are out of the scope of this deliverable, but they are supported by the framework.

In the SPEED-5G RRM framework, the grouped algorithms run sequentially and the final result is provided when all the algorithms terminate. Some of the algorithms may require more data from the system to do their calculations when they receive the start trigger. In order to maximize the usefulness of the final result of a group of algorithms, the framework provides an asynchronous buffered pipeline capability for grouping together a set of algorithms. This allows each algorithm to obtain all its required input, once the earlier algorithms in the sequence have provided their output.

Figure 3 shows how the framework operates. Note that the algorithm sequence is an example as different combinations may exist depending on the specific algorithms and use cases. The left side of the figure shows the sequential flow where the algorithms wait until a previous algorithm ends its calculation. The right-hand side shows the asynchronous mechanism where all the algorithms request data simultaneously when a new trigger is raised.

The new trigger for calculations arrives from the Message constructor. The RRM framework is responsible for notifying each algorithm in the affected group of algorithms that a new configuration output is required. Since the framework and the algorithms are decoupled, the framework provides the mechanisms that allow the algorithms to become registered for a specific set of received messages. The output structure provided by the group of algorithms is defined by the algorithm itself, while the framework is responsible for managing it properly. When an algorithm in the group is replaced, the adaptation layer is responsible for ensuring compatibility with the other algorithms and also with the messages which the RRM has to send or receive.



Result is that RRM provides nothing new. Process stopped

Figure 3: SPEED-5G RRM algorithms



On the asynchronous side, when the new structure arrives, the algorithms that require more data may request it through the Message constructor. The Message constructor sends the message response only to the algorithm which has made the request; but before the data arrives at this point, the algorithms will validate it. In case of the validation failure, the process is stopped and recorded into the Data Storage for further analysis.

The framework is able to work in Cloud Computing environments as well as in embedded equipment. For that reason, the algorithms inside this framework avoid storing data internally. The framework provides seamless access to data storage. As a result, the data to be stored is defined by the algorithm and managed by the framework. This is an important for the cRRM solution, as Cloud Computing requires stateless procedures.

The characteristics of the framework impose some constraints into the algorithms. For that reason, the algorithms deployed in this framework have to follow a specific design summarized in Figure 4.



Figure 4: Algorithm structure

The details of each functionality are:

- Validator: ensures that the content of the received message is within the range expected by the algorithm. The validators ensure the health of the system, avoiding unstable algorithms or bad states.
- Data Inspector: not all received data has, as a consequence, a new output value provided by the algorithms. For that reason, the new data received is validated in advance, in order to avoid unnecessary algorithmic or RRM operations. In addition, filters help avoid unnecessary message exchanges between RRM and external entities like the OAM or KPI Collector. If the filter does not discard the new input received, the algorithms have to provide a new output unless something is wrong in between. In addition, each algorithm knows which data has to be periodically received for ensuring the wellness of the network. The periodical request is also scheduled in the RRM framework by this module. The final implementation depends on how the other entities work.

- Data requestor: is the entity responsible for requesting more data when required. Once a new data input passes the filters and arrives at the algorithm modules, each one of them checks if they have all the required data. If not, each algorithm is capable of requesting them via Message constructor. See section 1.3.5 for more details.
- Calculator: provides the final solution. When the algorithm has all its required data, including the requested ones, it provides its final solution that is included in the output message frame defined by algorithms and managed by the framework.
- Message registrator: a lot of messages can arrive at the RRM but only a sub-set of them are intended for a given individual algorithm. This functionality is used by the algorithms to subscribe to the required ones, including periodic and asynchronous received messages. When a new message arrives, the RRM provides to each subscribed algorithm the content of this message.
- Internal data: functions to manage the information stored in the common data storage. Depending on the algorithm requirements, this data may be accessed only by the information owner. In case that any other algorithm requires access to this information, the framework facilitates the messaging between algorithms.

The Validators, Data inspectors, and Calculators are the common elements of an algorithm to guarantee an output while data request the periodic configuration request, the request reporter and the data manager are the new modules claim by SPEED-5G to use the novel RRM framework.

1.3.3 Requester

This functionality solicits the required algorithm information from the appropriate entity. The entity has to be configured and is the network operator the one which knows where the information is stored. On the other hand, it is the algorithm which knows which data is required. Therefore, an interface is required between the RRM and the OAM to finally configure the Requester.

The requester has an optional functionality for the periodic requests and it depends on how the other entities work and also based on the algorithm requirements.

The framework design is flexible enough to work in a heterogeneous network, being able to schedule a periodic report timer on an entity, subscribe itself to triggers, even establish the triggers, etc. In the case that external entities do not support these functionalities, it is the framework itself which creates the required requests.

The framework and algorithm decoupling splits also the functionalities. While the algorithm is responsible for indicating which requests are required, the framework is responsible for ensuring that request is transparent to the algorithm

1.3.4 Internal KPI collector

The RRM entity stores own KPIs and reports them to the OSS. The network manager, usually the operator, will define the required KPIs and are the algorithms the responsible for providing the information. The proposed framework provides the mechanisms to allow algorithm to report their KPIs and then, they are stored into the KPI Compiler.

The network manager configures the periodicity of KPI reporting for each report as multiple reports with different periodicity may be requested by the operator. That means the RRM framework has to create the connection, deal with the interface, create the reports and send the messages. In [Telecommunication management; Key Performance Indicators (KPI) for Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Definitions 32.450] we have an example of how the KPIs are defined by 3GPP. In this example, it is possible to see how the entity which has collected the KPI



information is responsible for distributing it. In our case, the algorithms are the entities which have the required information.

The following bullets show the required sequence before a KPI report is sent:

- The KPI report format is provided by the network operator.
- The network operator defines the required KPIs and how they are to be obtained.
- The algorithms are made ready to calculate the KPIs and send the results to the KPI Collector
- The KPI Collector picks up all the results and creates the appropriated message defined by the network operator
- The KPI Collector sends the message to the OSS defined by the network operator

Many KPIs may be calculated in parallel by different algorithms. The framework is responsible for ensuring the appropriate mechanism to receive the KPIs results, while the algorithm is responsible for providing the KPI results in time.

1.3.5 Message constructor

The Message constructor is the entity responsible for mapping the received data to internal and external format and forwarding it to the algorithms or to other entities respectively. This entity allocates the algorithms validators and filters. These algorithm functionalities are allocated here to avoid extra computational operations.

The main Message constructor functionalities are shown in Figure 5:

- Validators: validates the received data avoiding errors into the framework due to an incorrect or out of range value. This functionality is part of the algorithm as they are the only ones which know if the received data is valid or not for its purposes.
- Data inspector: avoids unnecessary process on the RRM framework by analysing at the process start if the new inputs are relevant or not. This is also an algorithm functionality.
- Message Creator: this entity maps internal messages into the output message and vice versa. For instance, when a cell sends a message to RRM, the received information may be useful for more than one algorithm. In this case, this functionality is responsible for filling each algorithm structure with the received data.





Figure 5: SW abstraction of the Message constructor

The Message constructor has to validate and filter new external data received, but filters may be avoided when the input is an internal module, or if it is data is received upon request.

The following steps are executed when a new message arrives from an external entity and the message has not been requested by RRM:

- 1. When an input from an external entity is received, the algorithm validator and data inspector, previously defined in section 1.3.2, evaluate the received information and discard it when it is wrong or the values do not require new input.
 - 1.1. In the case that all the responses of all the algorithms are rejected, the RRM framework stops the process and provides the answer if required.
 - 1.2. alternatively, the received data is forwarded to the Message Creator.
- 2. The Message Creator retrieves the appropriate message structure from the internal data in order to create the internal message that is provided to the Algorithms module. The specific message structure for each algorithm is defined by the algorithm or the group of algorithms themselves, as algorithms and framework are decoupled.
- 3. The Message Creator fills the retrieved structure and sends it to the Algorithms module. The main differences when a message arrives from an external entity are:
 - The received data does not go through the Data Inspector, as it has already been requested by the algorithm.
 - The Message Creator does not need to retrieve the message structure again. It simply fills the structure retrieved at the start.

Figure 5 also shows how this entity supports messages reported from internal modules. These messages are received when an algorithm provides the final solution, when it requires more data for its procedures, or when periodical requests issued by RRM.

The framework offers the possibility to record any message sent, for further analysis in the OAM. The same happens with the discarded input messages.



1.4 Interfaces

Every entity connected to RRM requires a specific interface based on as SCTP or TR-69 transport. All the required procedures to establish and maintain the interface connection are managed by the Message Dispatcher; such as, for instance, the required heartbeat to keep open an SCTP connection.

The Interface is also in charge of validating the received data, avoiding RRM malfunctions. It is important to note that the validation process over the Interfaces is different from the one that is done in the Message constructor. The Interface validates the data itself while the Message constructor validates the content of the message.

For the input procedures and once the data is validated, the interface sends the extracted message payload to the AL. On the other hand, for output procedures, the AL is the one which sends the message payload to be mapped into a standardized message.

1.5 Interfaces with the centralized RRM

One of the main SPEED-5G concepts is that the network intelligence is allocated in the RRM entity. For that reason, RRM is connected with most of the network elements. In the same way as the LTE X2 interface is used between cells to exchange information about handovers and for SON, the SPEED-5G project has defined a new set of interfaces, in which the information is grouped in a logical way depending on the task. A first version of these new interfaces are defined in tables in subsquent chapters of this deliverable; they will be further refined in D4.3.

For a better understanding of the RRM interfaces, the SPEED-5G architecture and the protocol stack defined in [1] is depicted in *Figure 6* and *Figure 7* below respectively.



Figure 6: SPEED-5G 5G Greenfield network architecture [1]





Figure 7: SPEED-5G Protocol Stack architecture [1]

1.5.1 5G RRC

The RRC layer is the one where the core and other cell connections end. For that reason, the RRC layer raises triggers and provides the required information regarding network control information. The interface between 5G RRC and RRM ensures the required backward compatibility with legacy RATs. The SPEED-5G project does not define any new functionality for RRC. Details provided in [2][3] are assumed by the project.

1.5.2 MAC

This is a key interface in SPEED-5G. All the messages required to make an efficient use of the spectrum rely on the communication between the new RRM algorithms proposed and the SPEED-5G MAC. Due to the importance of this communication, two different interfaces has been defined. One of them is dedicated to sending the channel sensed data while the other is used to send the configuration data. Full details on SPEED-5G MAC design can be found in [4].

1.5.3 KPI Collector

The KPI Collector, also called Cooperative Sensing GW, is basically a database that stores the KPIs reported by the cells. Some KPIs are standardized, but others may be proprietary. Therefore, the information stored in the KPI collector is dependent on the network operator requirements.

When the RRM algorithms require specific data which is not currently reported, the cells have to be updated/re-configured in order to report the required values. Note that how the data is obtained has to be perfectly defined. After the cells report the required values, it is essential to specify how the information will be retrieved from the KPI Collector by the RRM entity. This task is performed through the exchange of messages between the KPI Collector and RRM.

We provide an example with two different options for a better understanding. For example, when the RRM requires the throughput in an area during last 3 hours, there are different options that will define how the KPI Collector works. Case A: RRM asks a range of cells for inputs, over a specified period of time and KPI Collector reports each value during this period for each cell; Case B: RRM sends the same request as case A and KPI Collector sends the calculated value.



1.5.4 Spectrum Manager

The Spectrum Manager (SM) is the entity in charge of the shared spectrum usage across different geographical areas. In Europe, the Licensed Spectrum Access (LSA) framework is the initiative for regulating the shared spectrum so the SM requires an interface with it.

Based on the geographical location of the cell, the SM has to provide information about the available shared bands, including its load and how many operators are using it. This information will help the RRM to take its final decision about the spectrum selection or the spectrum aggregation. In case the SM is able to provide the pattern usage of a shared band, this information can be used by the RMM algorithms to provide a more accurate solution and also, forward the pattern to the MAC schedulers for an optimal spectrum usage. In short, the SM has to provide the following data:

- Operator spectrum ownership.
- Current spectrum allocations.
- Estimates of channel quality or occupancy at a geographical location.

The SPEED-5G project proposes a unique interface between the SM and the the centralized RRM. This solution avoids the need for (establishment & maintenance of multiple) interfaces between local dRRMs and the SM, avoiding the requirement for each node to establish a link with the SM.

Each operator is expected to have an SM that holds information on the spectrum (usage) that is licensed to that operator, and also on the unlicensed or lightly licensed bands that the operator is able to use. Virtual operators may share an SM with policies in place. In the case that each operator uses its own cells (no-infrastructure/RAN-sharing), the cRRM algorithms should be able to optimize the spectrum usage, but several cRRM instances may be needed to cover very large areas, and consequently, there is the need for an interface between cRRMs of the same operator for coordination purposes (considered out-of-scope). Multi-tenancy, where one or more cells are doing RAN sharing is where the same MAC scheduler is shared by more than one operator and therefore, it cannot receive configuration from more than one cRRM. For that reason, SPEED-5G suggests an interface exchange information between cRRM of different operators. Although more acceptable than sharing SMs, a potential problem with such an interface is that through it, the operators may have to share sensitive information about their own networks, including traffic load or the number of users per cell.

1.5.5 OAM/OSS

The RRM is a network entity that has to be managed by the OAM. The commonly-used interface protocol is TR-69, with different data models depending on the final equipment to be managed. The new interface between the OAM and RRM has to provide multiple information:

- Directions to reach other entities, such as, for instance, the IP direction of the KPI Collector or the Spectrum Manager.
- The set of configuration parameters required. The complete definition of these parameters is out of the scope of SPEED-5G.
- The framework stores its own KPI that will be reported to the KPI Collector or directly to the OSS in a period defined by the OAM.

1.5.6 Distributed RRM

This interface is used to transmit messages from dRRM and also from different cell entities to the cRRM. When the dRRM does not have any distributed algorithm, it works as a bridge between the cell and cRRM. In any case, the SPEED-5G proposal is able to support SON and the current state-of-the-art algorithms. If dRRM does not contain any algorithms, it only introduces delays and complexity into the network from the point of view of the cell. On the other hand, and considering cRRM, it has

to establish a single interface with the cell and it has therefore little internal functionality. This interface follows the same approach as S1, X2, or TR-69 interfaces, where a single interface is defined between two entities. Having a single interface facilitates maintenance, reduces complexity, and avoid extras configuration data.

The main function of the algorithms allocated to the dRRM is to provide output decision/response in almost real time. This decision is taken mostly based on the cell's own/local information and information received from neighbour cells. Since the information collection is reduced, the solution will deviate from the ideal one. When distributed solution crosses a defined threshold, the centralized algorithms can be triggered, to provide a more accurate solution and reset the distributed algorithms.

Taking this into consideration, the interface between centralized and distributed RRM has to contain messages will the following information:

- Message id, source, and target
- Any value required by the algorithms

The required information has to be grouped into messages and we can follow two different approaches. The one followed by Broadband Forum, the one followed by 3GPP or a hybrid solution. The Broadband Forum defines a single data model where a single message may contain all the information. On the other hand, 3GPP defines the messages based on the use case so, the total number of messages depends on the use cases required. Table 1 has a summary of the different approaches (cRRM or dRRM based) adopted by different RRM algorithms proposed in SPEED-5G.

1.5.7 Centralized RRM

An interface between cRRM is required in RAN-Sharing scenarios where a single cell supports more than one operator. From the point of view of RRM, the final configuration provided to the cell is a trade-off between cRRMs of different operators.

For the final message type decision, the same options are described in Table 2 to Table 3

1.6 Overview of RRM algorithms and functions

This section provides an overview of the functions targeted by the algorithms (load balancing, RAT selection, etc.) and introduces a table that maps functions and algorithms, identifies the location of the investigated functions (RRM or higher MAC), and classifies between centralized and distributed solutions.

Table 1 describe the algorithms, functionality and placement in system framework from different partners.

Algorithm target	Centralized or distributed?	Notes
Load balancing (steering of different traffic classes on the licensed and unlicensed band) & channel selection (for the unlicensed band if required) (red blocks in the MSC). Load balancing is done on a "long" time scale (say 80 to 100ms) and delivers a traffic steering decision on this time base. Between 2 decisions, several iterations of the "channel selection" algorithm are performed.	Our solution for the load balancing requires low latency BH, both dRMM and cRRM architecture are then feasible. It may work also in the case of BH with high latency but with reduced performance. Channel selection is located in the HMAC, which can be co-located with the RRM or not; we did not make any assumption with respect to this.	Load balancing requires average KPIs (cell capacity on licensed and unlicensed bands) and info on the buffer states. It may works also with outdated information but with reduced performance. Channel selection requires fast updates related to sensing and ACK/NACK.



Algorithm target	Centralized or distributed?	Notes
The algorithm targets the RAT spectrum selection and channel selection blocks (which communicate with the spectrum manager) and also emphasis will be given to the 3.5 GHz band specifically focused on SAS model (for unlicensed and lightly-licensed users).	Both cRRM and dRRM.	The algorithm requires up-to- date inputs but, it will also have our own mechanisms for coping with problematic data or in general with incomplete input.
Channel selection.	It is assumed to be executed at the eNodeB/AP of operators employing infrastructure sharing. Therefore, it is sitting in the dRRM.	Up-to-date information is necessary for the optimal operation of the algorithm. Stale input data will deteriorate its performance.
The proposed strategy mainly targets RAT selection and supports load balancing assisted by the network. According to the proposed framework, the network-side functional entities (i.e., policy designer and repository) would map to the cRRM of SPEED-5G, with a new entity (i.e., connection manager) introduced on the user equipment (UE) to implement the decision-making process. In term of intelligence, we can assume that, to some extent, we bypass the small-cell MAC by delegating the final decision to the UEs.	The proposed algorithmic solution is distributed: it runs directly on the UE dRRM (i.e., connection manager). The idea is to configure the UE to act on its own following a policy and set of controlling parameters that are set/adjusted on the network side.	For the short time-scale variability, the proposed solution does not need exact/much data from the network. The UE mainly relies on the local information it has access to. All that is needed from the network is that each small-cell broadcasts its load on its pilot channel/beacon. The UE cope with lack/incompleteness of information using fuzzy logic. For the longer time-scale variability, the cRRM could alter some parameters (e.g., the cost of a given RAT) to push some users towards a certain behavior e.g., perform a form of traffic steering to use some unloaded RATs during some periods of time.
It is to allocate radio resource allocation (carrier & power) considering different QoS requirements depending on traffic types. It is assumed that UEs are capable of accessing the cellular network and the WiFi network simultaneously to increase data rate and the eNB decides whether UEs will access the WiFi as well as the cellular network.	dRRM	It will need up-to-date inputs (i.e., current QoS level and network load). It will also use the information on the WiFi network status (i.e., RSSI and # of WiFi devices). Based on information on available spectrum & on up- to-date network's status, it will decide to allocate resource to users to maximize the user satisfaction level. Thus, providing accurate up-to-date inputs would be important.
The network side functional entities (resource allocation for co-existing LTE-U and WiFi networks to maximize throughput and hence minimize the interference.). A new entity named spectrum manager is needed that would track the interference and resource allocation on a different band, according to traffic conditions.	Centralized	Yes, need input on current KPI under consideration.

Table 1: Summary of algorithms

1.7 Demulator

The RRM (radio resource manager) is a crucial component of the SPEED-5G project, as it is responsible for carrying out the eDSA (enhanced Dynamic Spectrum Access) functions. As such, the

design proposed in earlier SPEED-5G deliverables needs to be tested in software before it is implemented in a full hardware testbed. This is why the concept of a demulator is proposed. A *demulator* is a piece of software which represents the combined functionality of a *demonstrator* and an *emulator*. Its intended function is to (1) demonstrate the working of the RRM algorithms in several scenarios closely connected to the SPEED-5G use-cases, and (2) incorporate emulation code for RRM functions which can eventually be used in the real system. The demulator is effectively a *reference model* for the RRM.

Recall that the cRRM is a high-level component of SPEED-5G, which operates at slow time scales, typically minutes or longer. One of its principal functions is to choose an appropriate RAT (radio access technology) for a specified service, upon request from a cell which has UEs requiring to be serviced. Because of the slow time-scale of operation, the RRM exists as a self-contained entity quite high in the network hierarchy and communicates with cells over TCP. Thus, the demulator runs as one or more processes on a single CPU with an allocated IP address. This design allows a distributed SPEED-5G architecture; in fact, the demulator may be sufficiently fully functional to act as a real RRM. The remainder of this section describes the software design of the demulator.

1.7.1 Software Design aspects

To ensure portability, and to enable high-level software design, the chosen language is Python (version 3.5 or higher).

The software makes use of the multiprocessing (mp) module of python, for maximum flexibility and portability over different thread implementations. The design is built upon several base classes, intended to be subclassed to cover various specific instances. The result is the following main classes:

- DuplexQueue: uses two SimplexQueues of the mp module, to allow duplex messagepassing between RRM components. Buffering is handled transparently. Each pair of components will install message handlers at each end of the DuplexQueue.
- 2. Logger: handles all system logging, writing tsv files for post-processing.
- 3. RRM_base: base class from which instances of RRM can be derived. Provides basic messaging, cell registration, and logging facilities, not explicit resource management functions.
- 4. MAC base: base class from which instances of the MAC can be derived.
- 5. Cell_base: base class from which instances of cells can be derived. Stores information on RATs of which the cell is capable.
- 6. UE base: base class from which instances of UEs can be derived.

1.8 Tables defining messages below the abstraction layer

Table 2 and Table 3 collect together all the low-level messages or primitives so far identified is being needed for the SPEED-5G RM framework. Additional messages might be added in later phases of the project, and empty slots are left for these in the tables. The exact mappings to the higher-level messages above the AL will be fully defined in D4.3.

Entity	Task	Interface Name	Description	Example
5G RRC	Reporting actions	C_CRRM-5GRRC SAP	Actions on RRC that requires approval has to be sent to	Starting a new service for an existing UE or a HO message arrives from a neighbour cell



Entity	Task	Interface Name	Description	Example
			RRM.	
	Configuration		Provides configuration data to RRC.	Deny all FTP services for a period of time
5G MAC	Sensing	S_HMAC_cRRM SAP	In order to optimize the spectrum usage, the sensed spectrum data is requested.	New device requests access and the current sensing results are required
JG WAC	Configuration	M_cRRM_Config SAP	(Re)configure the MAC to optimize the network performance.	A set of services are moved to another spectrum band
KPI Collector	Get information		Stores the system KPIs reported by every entity inside the system.	The load balancing wants to check if something is going wrong and KPI system is demanded
Spectrum Manager	Get information	C_SM-cRRM SAP	It has the Spectrum usage knowledge.	A current user or service has to be moved to another spectrum band
OAM	Configuration	C_5GOAM_cRRM	New RRM configuration set by the network operator. This interface also supports automatic RRM configuration update made by the algorithms which modify the operator configuration	Operators modifies spectrum allocation priorities
	Configuration	TBD	cRRM sends a new configuration to be applied	MAC layer has to modify FTP service from a licensed band to LAA
dRRM	Get information	TBD	Retrieve information from the remote entity	The cRRM requires instantaneous values from the admission control algorithm running in the remote cell
cRRM	Configuration	TBD	Used in combination of RAN-sharing	
5G RLC	Get information	TBD	Algorithms require the traffic buffer status before they provide the final	After congestion is detected, the load balancing algorithm requests the instantaneous RLC buffer status in order to provide the most suitable



Entity	Task	Interface Name	Description	Example
			solution	solution

Table 2: Summary of RRM interfaces defined in D4.1

1	Dne single message with definedEach time RRM has to send data, it will be all the encapsulated in a single		Minimize the load in the interface	All the algorithms have to check the received data.	
		message		The message is not sent until all the algorithms provides its parameters request/answer	
2	Message per use case with defined message fields	The message id is directly related to the information that it contains and there is a direct relation between the request and the response	Algorithms which received data is not relevant don't have to be activated	Due to the complexity of some algorithms, many messages may be required for a single algorithm	
			The response structure is directly got from the request	The message is not sent until all the algorithms provides its request/answer parameters	
3	Hybrid solution with a single message with defined messages	The message is internally structured by use case with a pre-defined structure. The	Minimize the load in the interface	The message is not sent until all the algorithms provides	
	field use case is easily getting by the RRM framework		The activated algorithms will be the ones which require this information	its parameters request/answer	
4	Hybrid solution with a single message without defined messages field	The message is internally structured by use case with a structure defined by the algorithms	Minimize the load in the interface and the algorithms activated will be the ones	Same use case may be repeated several times on the request with shared values	
			which require this information	RRM framework may send two different solutions for the same use case causing inconsistency on the target	

Table 3: Interface options between dRRM and cRRM

2 RRM algorithm 1: Efficient licensed-assisted access (LAA) operation in small cells, based on reinforcement learning

This algorithms is designed for operation in dense heterogeneous cellular networks with Licensed-Assisted Access (LAA) small cells, capable of operating in both licensed and unlicensed spectrum. To support broadband traffic services characterized by diverse Quality of Service (QoS) requirements, the proposed algorithm incorporates an optimized controller that makes efficient use of the available frequency resources, such that 1) QoS is maximized, 2) interference between cellular network and legacy WiFi using the unlicensed band is minimized, and 3) co-channel interference in the unlicensed band between nearby small cells is limited. Reducing the interference generated by neighbouring small cells on the unlicensed band is an important goal, in order to take full advantage of the available unlicensed band. Nevertheless, finding the optimal association between available channels and small cells remains a combinatorial problem, which cannot be solved with limited complexity. Therefore, a distributed solution, based on reinforcement learning is proposed, where each small cell can autonomously learn "by interaction", what is the most appropriate unlicensed channel to select.



Figure 8: The heterogeneous network under investigation.

2.1 Relation to Golden Nugget

In the following, we highlight the relation between the proposed solution for licensed assisted access and the SPEED-5G golden nuggets.

- Develop hierarchical (blending distributed and centralised) management of ultra-dense multi-RAT and multiband networks
 - In this proposal, the traffic steering controller located in the RRM can be a centralized function while the channel selection controller is located in the Higher MAC, which can be distributed
- Capacity → Through small cells and efficient resource allocation
 - The goal of this study is to increase the small cell capacity through an effective utilization of unlicensed bands
- ► Scalability → Through distributed management
 - The distributed channel selection function can provide higher flexibility and system responsiveness

- ► Stability → Through machine learning and assuming certain levels of traffic for specific timeframes
 - The proposed solution is based on the multi-armed framework, which is a machine learning tool

2.2 Assumptions and system model

We investigate downlink communications in dense heterogeneous networks with UEs and small cells having LAA capabilities. More specifically, small cells can be deployed either in co-channel mode, using the same band as the macrocells (e.g., 2GHz), or in dedicated channel mode (at 3.5 GHz). In both cases, the available licensed bandwidth is assumed to be 10MHz. In addition, small cells can operate in the 5GHz unlicensed band where eight channels of 20MHz are available for shared access. All the small cells are assumed to be time-synchronized and unaware of the unlicensed band occupancy statistics (probability of being vacant, transition probability from vacant to busy, etc.). Other system parameters are in line with the 3GPP specifications [5].

2.3 Algorithm Description

At the cRRM, the traffic steering function has the role of continuously controlling whether the small cell load and the QoS requirements can be satisfied by using only the licensed band or the access to the unlicensed band is needed. This requires monitoring the status of the QoS of active traffic classes and the capacity of the associated users

cRRM Monitor
$$\sum_{i} \overline{R}_{i,j} \ge \sum_{i} R_{i,j} \quad \forall j \in TC, i \in UE_{j},$$
 (1)

where $\bar{R}_{i,j}$ represents the estimated data rate at the UE *i* characterized by the traffic type *j* and $R_{i,j}$ its actual rate requirement.

When this requirement is not satisfied for all the traffic classes, the most demanding traffic class is steered on the unlicensed band. Then, to select the appropriated channel to access, we use a Bayes Upper Confidence Bound algorithm (Bayes UCB) [6], which enables us to estimate the channel statistics and to take the best decision accordingly. In particular, each timeslot t, a small cell selects a channel $k \in Ch\{1, ..., N_{Ch}\}$, implement the LBT function, and if it is available the related resources are scheduled to the UEs belonging to the traffic class indicated by the cRRM. In addition, when multiple small cells access to the same idle channel, packet lost may occur due to the co channel interference. Let $r_{k,t}$ be the instantaneous reward associated to the channel k at the time slot t computed as

$$r_{k,t} = \begin{cases} 0 \ if \ \text{WiFi is detected or packet loss occurs} \\ 1 \ else \end{cases}$$
(2)

Then, the aggregated reward for the channel k over the horizon T = 1, ..., t is computed as $R_{k,t} = R_{k,t-1} + r_{k,t}$. Then the objective of the controller is to find a strategy that maximizes $\sum_{k}^{Ch} \sum_{t}^{T} R_{k,t}$ or equivalently the regret with respect to the ideal policy, which knows the channel statistics and distributes the small cells such that the experienced interference is minimized (that is, orthogonal allocation). The Bayes UCB is asymptotically optimal and characterized by logarithmic regret with respect the ideal policy, which is the best achievable performance by a solution that does not use *a-priori* information on the channel statistics.

To deal with the small cell interference, the classic Bayes UCB is modified by considering that each small cell n maintains a rank value v_n [7], which is updated each time a collision occurs, such that nearby small cells do not select the same channel amongst the available ones.



2.3.1 Algorithm pseudo-code

The overall Ba	ayes UCB algorithm, which each small cell n implements at each timeslot t , is as follows:
Inputs: T, N _C	$h, v_n, R_{k,t-1}, N_{k,t-1}$
Outputs: v_n ,	$R_{k,t}, N_{k,t}, S_{k,t}$
1.	for $t = 1$ to T
2.	if any $(N_{:,t-1} = = 0)$
3.	$S_{n,t} = i \text{ s.t. } N_{i,t-1} = 0$
4.	Else
5.	for $k = 1$ to N_{Ch}
6.	$q_{i,t} = Q\left(1 - \frac{1}{t}, Beta(R_{i,t} + 1, N_{i,t} - R_{i,t} + 1)\right)$
7.	End
8.	$S_{n,t} = i$ s.t. $q_{i,t}$ is the v_n^{th} larger value of $q_{:,t}$
9.	End
10.	if $S_{n,t}$ is vacant
11.	$R_{i,t} = R_{i,t-1} + 1$
12.	End
13.	$N_{i,t} = N_{i,t-1} + 1$
14.	End

where $N_{i,t}$ is the number of times the channel i has been selected up to the time slot t.

2.3.2 Inputs and Outputs

The Input and output variables which the proposed hierarchical machine-learning algorithm will use and produce are illustrated in *Table 4* and *Table 5*

Parameter name/ID	Description	source CRRM INTERNAL block	INTERNAL interface name/ID	source EXTERNAL block	EXTERNAL interface name/ID
CQI	CQI-Update on Licensed and Unlicensed Bands			РНҮ	M_PHY_HMAC SAP
Sensing	Outcome of the LBT scheme (channel busy/idle)			РНҮ	M_PHY_HMAC SAP
Cap_Lic Cap_UnLic	Cell Capacity on Licensed and Unlicensed Bands			HMAC	M_HMAC_cRRM SAP
Buff_state	Service load, PER and HoL delay per queue			RLC	C_RLC_cRRM SAP

Table 4: List of INPUT parameters used by the efficient LAA operation in small cells.



Parameter name/ID	Description	destin ation CRRM INTER NAL block	INTERNA L interface name/ID	destination EXTERNAL block	EXTERNAL interface name/ID
LB_O	Traffic steering on lic. And unlic. Bands			dRRM/ HMAC	C_CRRM_HMAC SAP
Conf_Sch	Map Lic. And Unlic. scheduler with active queues			dRRM/ HMAC	C_HMAC_LMAC SAP
Conf_Un_Chan	Map Unlic. Channel with Unlic. Scheduler			dRRM/ HMAC	C_HMAC_LMAC SAP

Table 5: List of OUTPUT parameters/values provided by the efficient LAA operation in small cells

2.3.2.1 Mapping of proposed algorithms to the cRRM/MAC blocks

The algorithm operation is line with the protocol architecture proposed by SPEED-5G (see Figure 9). Specifically, it is assumed that the CRRM block observes the RLC buffer status and periodically decide whether to offload the licensed spectrum and how to *steer* the active streams (i.e., traffic class) across the licensed and unlicensed band. When the offloading decision is taken, this input is transferred by the cRRM to the dRRM /higher MAC, where the MAC controller is in charge of selecting the appropriate unlicensed channel to sense according to the history of the unlicensed channels (i.e., spectrum availability and measured SINR).

Then, the listen-before-talk (LBT) procedure is implemented [8] and if the selected channel is sensed as available, the secondary carrier is configured and the related stream is scheduled by the Lower MAC functions. On the contrary, when the channel is sensed as busy, the transmission is denied, and the MAC controller adjusts the cRRM offloading decision accordingly, i.e., during the next 10 Transmission Time Intervals (TTIs), all data traffic will only be transmitted in the licensed band. To conclude, we assume that the Lower MAC operates at the TTI granularity (1ms), the Higher MAC controller works at a frequency of 10 TTIs, and the cRRM in an even slower time scale.

The message sequence chart providing the details of the exchanges across the different logical entities is described in Figure 10.



Figure 9: Mapping of the efficient LAA operation in small cells to CRRM/MAC blocks.





Figure 10: Message sequence chart for the efficient LAA operation in small cells, based on reinforcement learning.

2.3.3 Simulation assumptions and parameters

Table 6 shows the main simulation assumptions used to evaluate the proposed solution. Other system parameters are in line with the 3GPP specifications [5].

Since the focus is on small cell deployments with LAA capability, we only report on the performance of the small cell users. As a preliminary evaluation, the proposed algorithm results are compared with an access scheme where the LAA channel is selected according to a random policy. Then, we evaluate the successful access probability (which account for both WiFi collision and LAA small cell interference), the user experienced latency and throughput.

Parameters	Values	Parameters	Values
ISD	500 m	# of SCs	4 per Macro sector
# of Macro	19 three-sectorized	# of UEs	30 (2/3 of the UEs are randomly located
eNB			in the hot-spots)
Available	8 x 20 MHz @ 5 GHz	UE Traffic	50% NRTV (rate 512 Kbps, latency 100
spectrum	1 x 10 MHz @ 3.5 GHz	type	ms latency)
	-		30% FTP (rate 500 Kbit, average reading
			time 0.1 s, and latency 300 ms)
			20% CBR (rate 1Mbps)
LAA channel	(0.1; 0.2; 0.2; 0.8; 0.6; 0.4	4; 0.1; 0.3)	
busy			
probability			

Table 6: Simulation assumptions and parameters of proposed algorithm

2.3.4 Performance evaluation

The obtained results confirm the effectiveness of the multi-armed bandit solution. In fact, Figure 11 shows that, after convergence, the proposed algorithm provides 25 percent higher probability of successful access to the LAA channel, compared to random selection policy.





Figure 11: Probability of successful access in the LAA band with the Multi-armed Bandit and a random access scheme.

Such a higher probability leads to lower experienced latency (see Figure 12), especially for the cell-edge user and for traffic with lower latency constraints (FTP and CBR).



Figure 12: Cumulative Distribution Function of the user latency with the Multi-armed Bandit and a random access scheme.

Finally, we also observe a higher user throughput (see Figure 13), particularly related to the FTP traffic, for the cell edge users.



Figure 13: Cumulative distribution function of the user throughput with the Multi-armed Bandit and a random access scheme.

3 RRM algorithm 2: RAT/spectrum/channel selection based on hierarchical machine learning

The algorithm is intended for RAT/spectrum/channel selection based on machine learning and is designed for operation over the 3.5 GHz lightly-licensed band for achieving better performance, especially in dense and congested 5G environments. According to various studies, this band may be a good option for ISM (when not occupied by incumbents) in order to boost capacity, and used especially by small cells with relatively limited transmission power and range. Our work takes into account the fact that we have a pool of bands and various licensing schemes (licensed/unlicensed/lightly-licensed) and we need to fulfil certain traffic requirements. As a result, we are looking for a solution for effectively addressing this situation and at the same time create value for the stakeholders involved [9].

3.1 Relation to Golden Nugget

This part of the work is motivated by the fact that the project proceeds also to the development of a hierarchical (that is, blending distributed and centralised) management of ultra-dense multi-RAT and multiband networks, which is one of the golden nuggets that have been proposed by SPEED-5G in the 5G-PPP association. In SPEED-5G, centralized management is used as a baseline, as shown in Figure 14a) which can be expanded with distributed management by moving management decisions related to RAT/spectrum/channel selection closer to the node level as depicted in Figure 14b. Specifically, the proposed algorithm will initially run in a distributed manner in order to limit the excessive signalling of centralised solutions in dense environments. However, in the case where the distributed approach does not provide satisfactory solutions, then a centralised approach could be used.



Figure 14: Centralized vs. distributed management.

Based on the increased demand for sub-6 GHz spectrum needed for future wireless networks due to the exponential growth in wireless data traffic, changes must be done in order to obtain quality of service and capacity expansion. Therefore, regulatory bodies are increasingly pursuing policy innovations based on the paradigm of shared spectrum, which allows spectrum bands that are underutilized by primary owners to be exploited opportunistically by secondary devices. Specifically, our algorithmic solutions will be focused to the Spectrum Access System (SAS) in the 3.5 GHz band, which consists of a three-tier model.


- SAS model (three layers).
 - Incumbent Access
 - Priority Access Licenses (PAL)
 - Lightly-Licensed
 - General Authorized Access (GAA)
 - Unlicensed



Figure 15: SAS three-tier Model.

Figure 15 depicts the SAS hierarchical model of the three tiers: Incumbent, PAL and GAA usage priority. The higher users are in the pyramid are able to utilize a channel over the user that is at a lower level. Also the two levels that are presented with a dashed line forming an isosceles trapezoid, shows the association of these two users with the SAS system. In particular, PAL and GAA users are controlled by the SAS system and thus must register and check all of their operations in order to provide a secure and interference-free environment to higher-tier users.

Related to machine-learning aspects, supervised and unsupervised learning could be considered for RAT/spectrum/channel selection. Supervised learning could be achieved, for example, through channel segregation techniques while unsupervised learning could be performed through Self-Organizing Maps (SOM) which can show available portions of spectrum such as heat maps.

3.2 Assumptions and system model

The basic proposed rules required for the SAS model that must be introduced for the usage in the 3.5 GHz band are:

Incumbent Access

- Citizens Broadband Radio Service for shared wireless broadband use of the 3550-3700 MHz band.
- Incumbent Access users include authorized federal and grandfathered Fixed Satellite Service users currently operating in the 3.5 GHz Band.
- These users will be protected from harmful interference from Priority Access and General Authorized Access users.

Priority-Access Licenses (PALs)

- Priority Access Licenses (PALs) will use competitive bidding within the 3550-3650 MHz portion of the band as illustrated in Figure 16 (blue-colored spectrum).
- Each PAL will use a 10 MHz channel in a single census tract for three-years (have to acquire a specific license).
- Up to seven total PALs may be assigned in any given census tract with up to four PALs going to any single applicant.
- Applicants may acquire up to two-consecutive PAL terms in any given license area during the first auction.

General Authorized Access (GAA)

- General Authorized Access users are permitted to use any portion of the three.
- 550-3700 MHz band not assigned to a higher tier user.
- May also operate opportunistically on unused Priority Access channels 3650-3700 MHz as illustrated in Figure 16 (Green-colored spectrum).



- GAA users are permitted to use 80 MHz of all the available and not assigned to any higher tier 3.5 GHz band.
- GAA do not have to obtain an individual spectrum license.

As can be seen, the regulations and specifications listed above for the SAS model on the 3.5 GHz are quite complicated, and are necessary in order for the model to work properly and thus should be introduced to the simulator. This is essential for the evaluation of the proposed solution and of significant importance so that the results will be as close to the real-life implementation as possible.



Figure 16: 3.5 GHz spectrum allocation for each tier of the SAS model.

3.3 Algorithm Description

3.3.1 Algorithm flowchart

The proposed solution of channel selection based on the 3.5 GHz SAS model is:

- GAA Users could use learning mechanism to get:
 - Channel usage of PAL Users
 - Channel usage of neighboring GAA Users





Figure 17: Flowchart of algorithm with learning capabilities as an option

For example, band, duration of usage, recurrence of usage, location, in order to be able to select and change to a specific cell/channel faster and more reliably by predicting the utilization of the 3.5 GHz band of the Incumbent and Priority Access Licenses users. Figure 17 illustrates a flowchart of the proposed algorithm and the procedure of learning and predicting the conditions of the channels from the view of the GAA users that are the most complex of the three tiers to assign a channel, but in general the same aspects apply to the PAL licensed users.

Description of flowchart:

The presented flowchart illustrates the steps of the algorithmic solution for utilizing the 3.5 GHz band in our system that focus on the SAS three-tier lever mechanism that is addressing how the different licensed users transmit at the specific band. Specifically, *Figure 17* highlights how the proposed algorithm will work with learning capabilities as an option when employed jointly with the SAS system. By involving the block of learning algorithm the system will be able to acquire information about the conditions (e.g., throughput, SINR, etc.) of the channels at a future time that the SAS system is not capable of providing.

Learning-based channel selection:

As illustrated in the *Figure 17*, the "Learning-based channel selection" box introduces a learning mechanism which will be able to provide information and predictions about the characteristics of the channel and more specifically the quality of each channel. In general, by applying statistical learning techniques will be easier to automatically identify patterns in data that can be used to make more accurate predictions. At this part of the main algorithm, has to be determined whether a channel is good or bad. In machine learning terms, a classification task will be introduced to categorize data points and for example, based on the maximum transmitting throughput, the channel that reaches high throughput should be classified as the best channel and the one to be chosen. The learning method as mentioned will find patterns and identify boundaries at given data. Acquired data can be split into two parts, the training data which will be used to train the model of



the algorithm and the test data which will be used to test the model's performance on that new data that have never been applied to the algorithm before.

General algorithm analysis:

The data which will be acquired from the system could be info of the three different licensing tiers on channel usage, duration and even recurrence of usage and in general acquisition of information for the availability and the conditions of each channel. With that in mind would be possible to locate the best channel that should be utilized each time. The information collected will be for each specific location in order for the algorithm to provide the cRRM and dRRM with the appropriate knowledge of the system when is required.

In addition, after the step that a 3.5 GHz channel is utilized (marked with a yellow arrow as "Output") that algorithm will keep checking for any information about the channel that is given to a specific tier. In particular, the PAL and GAA users may receive instructions to change channel (or even band if no channels are available) whenever a higher tier user needs to use the specific channel. By leveraging on the basic SAS principles the algorithm will be able to acquire knowledge about the channels utilization and specifically, if the band is occupied by Incumbent, PAL or GAA. Most importantly the SAS is not a real-time scheduler and this is why the proposed algorithm is needed, and will provide the appropriate efficiency and effectiveness of faster and better selection of the channels without any implications in proper functioning of the system.

3.3.2 Inputs & Outputs

The Input and output variables which the proposed hierarchical machine-learning algorithm will use and produce are illustrated at Table 7 and Table 8 below:

		Inputs			
Parameter name/ID	Description	Provided by CRRM INTERNAL block	INTERNAL interface name/ID	Provided by an EXTERNAL block	EXTERNAL interface name/ID
Ν	Nodes	-	-	Yes	C_5G-X2AP
В	Spectrum Bands	-	-	Yes	C_SM-cRRM SAP
С	Channels	-	-	Yes	C_SM-cRRM SAP
UEs	User-equipped devices (CBSD devices of PAL and GAA users that also implies the License scheme of the users)	-	-	Yes	-

Table 7: Input parameters of proposed algorithm

	Outputs							
Parameter name/ID	Description	Supplied to CRRM INTERNAL block	INTERNAL interface name/ID	Supplied to an EXTERNAL block	EXTERNAL interface name/ID			
Selected Channels	The channel that each node is assigned to in order to transmit	-	-	Yes	M_cRRM_ConfigSAP			
Throughput		-	-	Yes	KPI collector			

Table 8: Output parameters of proposed algorithm



3.3.2.1 Mapping of proposed algorithms to the cRRM/MAC blocks

In Figure 18 the mapping of the algorithm is illustrated. Specifically, there can be seen the connection of the mechanism to the RRM functional blocks, RAT/Spectrum Selection and Channel Selection. Those blocks, of course, are connected with the spectrum manager, which also provides the appropriate data of the spectrum utilization as an input to the system. Also, when running in a more centralized way the spectrum manager will provide our algorithm with the operator spectrum ownership, the current spectrum allocations and even estimations of channel quality or occupancy at a geographical location. In addition, various data about the cells will be produced at the 5G-CELL block, and supplied to the RRM blocks by the utilization of the C_5G-X2AP interface as shown. Likewise, the algorithm will be able to connect with the KPI collector where it can retrieve or send information about a specific cell or a range of cells of an area. Finally, the RRM algorithm will be able to exchange information through a two way communication with the 5G MAC layer. More specifically two interfaces, the S_HMAC_cRRMSAP will provide inputs to the RRM blocks that will be used by the algorithm and the M_cRRM_ConfigSAP interface will support the information provided by the RRM to the MAC interface for configuration or reconfiguration when appropriate and required/requested by the algorithm.

Figure 19 is a chart of the messaging sequence of the proposed mechanism. Initially, the algorithm will first run in a distributed way, in every cell inside the dRRM. From there, information about the availability of RAT/spectrum/channels will be gathered and the learning mechanism will be introduced in order to acquire information about the cell's users, utilized licence schemes and neighboring cells frequency assignments, etc. After a successful selection or even prediction of a particular channel has been completed, all the appropriate information and specifications will be requested and received to and from the MAC layer. At the MAC layer scheduling and inter-RAT coordination mechanisms will be enabled and run if necessary and send information to the dRRM layer. Then, the physical layer will be reconfigured to the new RAT, band, and channel. Finally, it should be mentioned that the cRRM will be invoked whenever the dRRM algorithm does not achieve the desired KPIs, or in general the thresholds required by the system for optimal results. Hence, the dRRM sends the available information from all cells to the cRRM where a centralized channel selection will run in order to produce the best possible solution. At that time, the MAC is called to perform the scheduling and inter-RAT coordination, sending new configuration instruction to all dRRM.



Figure 18: Mapping to RRM functional blocks as defined in SPEED-5G D4.1.





Figure 19: Message sequence chart for RAT/spectrum/channel selection based on hierarchical machinelearning.

3.3.3 Performance measures and KPIs

The performance measures and KPIs that will be used in the implementation of the proposed algorithm are introduced and explained in the below table (Table 9).

	Requirements			
КРІ	Massive IoT communications	Evolved Mobile Broadband		
User Experienced Data	From tens to hundreds of Khos	DL: 300 Mbps		
Rate	From tens to numbreus of Kbps	UL: 50 Mbps		
	On demand	On demand,		
Mobility	On demand	0-100 km/h		
Connection Density	Up to thousand devices/km ²	200-2500 /km ²		

Table 9: Performance measurements and KPIs of proposed algorithm

3.3.4 Simulation assumptions and parameters

For the simulation, seven to nineteen 3-sector macro base stations (as 3GPP requires at their scenarios) will be utilized, and up to tens of small cells for each macro and run multiple scenarios and tests cases.

Simulation parameters are presented at the table below (Table 10):



Nodes	Location	Central frequency	Maximum Tx power	Bandwidth
Small	Outdoor	3.5 GHz	2x27 dBm	10 MHz
Macro	Outdoor	2 GHz	3X2X46 dBm	20 MHz

Parameters	Values
Physical layer parameters	Channel Bandwidth: 10MHz,
	Subframe length: 1ms,
	Number of resource blocks: 50,
Resource block format	Number of subcarriers per resource block: 12,
	Number of symbols per resource block: 7,
	Subcarrier spacing: 15kHz
Modulation and Coding Schemes	64QAM 4/5 (DL) & 16QAM 3/4 (UL)
Path loss model	2 GHz: L = 128.1 + 37.6log10(R), R in km
	3.5GHz small cells: L= 140.7+36.7log10(R), R in km
Traffic model	FTP
Simulation time	60s

Table 10: Simulation assumptions and parameters of proposed algorithm

Table 11: General simulation assumptions and parameters

3.3.5 Performance Evaluation

In order to evaluate the proposed algorithm for hierarchical learning in a heterogeneous system as descripted, a proprietary simulator shall be utilized with all the above parameters implemented. Therefore different scenarios and test cases will be introduced to the simulation environment and will be further described and analysed in details. In general, we will consider a sparse to dense environment based on density of Users and mobile devices. Additionally we will experiment with various triggering situations about the usability of channels from different licensed users in order to obtain an even broader knowledge of the algorithm capabilities, and overall performance for the specific network environment that will be introduced to the simulator.

Initially, a more high level approach of our algorithm is deployed to the system-level simulator that utilises a number of macro and small cells at 2 GHz and 3.5 GHz bands respectively with a number of different channels that every user is assigned to. Specifically, the parameters imported to the simulator are, 19 macro base stations (BS) each with three cells and also three small cells per BS, giving us a total of 114 cells throughout the network. In addition we have utilized 4 channels at 20 MHz bandwidth for every cell. The small cells are located close to the center of the macro cells and are working at the 3.5 GHz band in contrary to the macro cells that are working on 2 GHz, giving us a heterogeneous environment for our scenarios.

Figure 20 (a) indicates the results from our experimentation. The bar chart illustrates the cases of Table 12 and specifically the average duration in time that is needed for a successful completion of a session. On average the SINR-based approach has good results and we only see highest session durations when the network system is fully loaded with the bigger file size at 32 Mbit, meaning that the mechanism can be used most of the times offering satisfying results.







Figure 20: a) Average session duration, b) Average cell-edge throughput (5%)

Index	Users	Session/Day/User	Packet size
1	1140	720	16 Mbit
2	2280	720	16 Mbit
3	4560	720	16 Mbit
4	1140	1440	16 Mbit
5	2280	1440	16 Mbit
6	4560	1440	16 Mbit
7	1140	720	32 Mbit
8	2280	720	32 Mbit
9	4560	720	32 Mbit
10	1140	1440	32 Mbit
11	2280	1440	32 Mbit
12	4560	1440	32 Mbit

Table 12: Tested scenario cases

At Figure 20 (b) can be seen the cell-edge throughput of the network generated by the tool. This graph summarizes the main idea of the network capabilities when an interference-aware mechanism is introduced, which manages in a better way the channels that each UE is assigned. So as a conclusion from these test cases for the throughput capabilities is that the SINR mechanism can produce good results on average and can be introduced at the main algorithm in order to better decide how to assign the most appropriate channel to a particular user (or even user category).

These are preliminary results and further, updated performance evaluation of our approach will be available in D4.3.



4 RRM algorithm 3: Radio resource allocation with aggregation for mixed traffic in a WiFi coexisted heterogeneous network

While mobile data traffic is expected to grow significantly, almost 70% of mobile data traffic will be from video by 2021, which means 55% growth annually [10]. In order to support the growth of traffic, the use of multiple spectrum bands by carrier aggregation (CA) has been considered [11]. Along with licensed spectrum, the use of unlicensed spectrum has been investigated, i.e., LTE for Unlicensed (LTE-U), Licensed-Assisted Access (LAA)-LTE and LTE-WiFi Aggregation (LWA)[12]. Especially, LWA is a link aggregation technology that combines two different radio access technologies (RATs) – LTE and WiFi – while the conventional carrier aggregation (CA) in LTE-A combines multiple LTE carriers. LWA enables a device to use both LTE and WiFi networks simultaneously, so it can significantly enhance data rate by combining two networks' best achievable rates [13]. Our work considers the resource allocation with aggregation to support different level of quality of service (QoS) of different traffic types in the cellular network where the WiFi network coexists.

4.1 Relation to Golden Nugget

This work is motivated by the fact that the WiFi network can be utilized with the cellular network to increase the data rate of UEs capable of dual connectivity with LWA model. LWA emerges as an alternative to LTE-U/LAA which requires new 5GHz LTE-enabled device and small cells for unlicensed band uses. For LTE traffic transmission, LWA uses unlicensed bands just like LTE-A/LAA, but the transmission is made through WiFi, unlike LTE-U/LAA. This indicates LWA does not need new LTE-enabled 5GHz HW, and it can transmit LTE traffic through WiFi APs connected to LWA base station. Thus, LWA uses LTE on LTE bands and WiFi on WiFi bands. This eliminates potential fairness or regulation issues on WiFi bands. Considering widespread availability of WiFi and the fact that it operates over unlicensed bands, using the WiFi network can significantly contribute to increasing the data rate of cellular users.



Figure 21: An overview of LTE-WiFi Aggregation (LWA) [14]

In such multi-RAT heterogeneous networks, optimal resource allocation becomes an important issue since the different RAT have different characteristics along with different operating spectrum bands. Here, we investigate the efficient resource allocation considering the nature of the network for UEs of heterogeneous traffic types. Thus, this work could support the implementation of Enhanced Dynamic Spectrum Access (eDSA) by contributing to increasing the efficiency in the use of multiple spectrum bands while providing better QoS of different traffic UEs.



4.2 Assumptions and system model

We consider downlink transmission in heterogeneous networks as depicted in Figure 22. The system consists of a base station (eNB), *K* users (UEs) with cellular (e.g., LTE) and non-cellular (e.g., WiFi) air interfaces. *N* carriers in multiple licensed bands are available. There are a WiFi AP and *W* wireless nodes (WNs) with only WiFi air interface.



Figure 22: Support for multiple users of heterogeneous traffic types with the use of carriers in multiple licensed bands and WiFi network

By using carrier aggregation, multiple carriers in different bands can be integrated for the transmission channel. Then, the aggregated channel can include a primary carrier (PC) and multiple secondary carriers (SCs). While the PC provides a reliable connection, SCs can provide higher data rates and more capacity [15]. In addition, by LWA, the link of the WiFi network can be aggregated to improve data rates.

The data traffic is classified into two categories, inelastic (real-time traffic) and elastic traffic (besteffort traffic). While different traffic types require different levels of QoS, the approach of utility functions incorporating required data rates is considered to express the satisfaction level of required QoS. The utility function of UE *i*, $U_i(r_i)$ is represented by sigmoidal-like function or logarithmic functions depending on traffic types. These functions have the following properties: 1) $U_i(0) = 0$ and $U_i(r_i)$ is increasing function of r_i , 2) $U_i(r_i)$ is twice continuously differentiable in r_i [16].

For UEs of inelastic traffic, we use the normalized sigmoidal-like utility function, that is

$$U_i(r_i) = c_i \left(\frac{1}{1 + e^{-a_i(r_i - b_i)}} - d_i \right),$$
(3)

where $c_i = (1 + e^{a_i b_i})/e^{a_i b_i}$ and $d_i = 1/e^{a_i b_i}$ so satisfying $U_i(0) = 0$ and $U_i(\infty) = 1$.

For UEs of elastic traffic, the normalized logarithmic utility function is utilized and represented as

$$U_{i}(r_{i}) = \frac{\log(1+k_{i}r_{i})}{\log(1+k_{i}r_{i}^{\max})}$$
(4)

where r_i^{max} is the rate for UE *i* to achieve 100% utilization and k_i is the slope of the curve that varies based on the user application (i.e., the increasing rate of utility percentage with the allocated rate r_i). Figure 23 shows curves of various utility functions of different parameters. Considering that r_i can be obtained by multiple carriers (i.e., primary and secondary carriers), the utility functions can be represented as multi-variable functions.





Figure 23: Curves of utility functions with different parameters for heterogeneous traffic (blue colour for inelastic traffic and purples colour for elastic traffic)

4.3 Algorithm description

4.3.1 Algorithm flowchart

The proposed resource allocation and aggregation algorithm are illustrated in Figure 24.

For given carriers in multiple bands, carriers of the same number of UEs are selected from lower bands for the primary carrier allocation. By using the different utility functions, a single primary carrier is determined to each UE. For UEs requiring more data rates, we have two options: 1) using WiFi network, 2) allocating secondary carriers in licensed bands. While exploitation of WiFi network can increase the data rate without using licensed band resource, the gain would not be promising when the WiFi network traffic load is very high. Thus, after estimating the expected data rate achievable by the WiFi network, it is decided whether the WiFi network will be utilized or not in addition to the effective number of UEs. To choose proper UEs for WiFi, we introduce UEs' access index considering the UE's moving speed, the received signal strength power, and the traffic types. While the access index indicates the appropriate level to access the WiFi network, UEs having higher index values can be chosen for WiFi to increase the data rate. For UEs not chosen for WiFi, secondary carrier allocation is carried out from the licensed bands carried out from the licensed bands.



Figure 24: The flowchart of the proposed resource allocation algorithm

While the high-level procedure of the proposed algorithm is illustrated in Figure 24, we present details of each step for the proposed resource allocation and aggregation.

Resource allocation for Primary and Secondary Carriers

We consider the overall objective of the resource allocation to maximize the level of QoS guarantee to each UE. While the resource allocation is carried out with two steps: 1) primary carrier allocation and 2) secondary carrier allocation, each step shares the common approach. The only difference between two steps is that a single carrier is allocated for the primary carrier while multiple carriers can be allocated for secondary carriers.

By using aforementioned utility functions, the resource allocation problem is formulated as follows.

$$\max_{\alpha_{ij}} \prod_{i=1}^{K} U_i(\sum_{j=1}^{N} \alpha_{ij} r_{ij})$$
(5)

Subject to

(C1)
$$\sum_{i} \alpha_{ij} \le 1$$
, $\forall j$, (6)

(C2)
$$\alpha_{ij} = \{0, 1\}, \quad \forall i, j,$$
 (7)

(C3)
$$\sum_{i} \alpha_{ij} r_{ij} \ge R_i^{\min}$$
, $\forall i$, (8)

The objective (5), the function of utility proportional fair (PF), is to allocate resources for each user to maximize the total system utility while ensuring proportional fairness between utilities. The function ensures the non-zero resource allocation for all UEs. Since $\arg \max \prod U_i(x)$ is equivalent to $\arg \max \sum \log[U_i(x)]$, we will use $\sum \overline{U}_i(x) = \sum \log[U_i(x)]$ instead of $\prod U_i(x) \cdot \alpha_{ij}$ is the carrier allocation indicator, i.e., $\alpha_{ij} = 1$ indicates that carrier *j* is allocated to UE *i* and $\alpha_{ij} = 0$, otherwise. Since each carrier is exclusively allocated to one UE, the constraint in (C1) is imposed With Shannon capacity, the data rate is expressed as $r_{ij} = w_j \log_2(1 + r_{ij}p_{ij})$ where w_j is the bandwidth of carrier *j*. For UE *i*, the minimum rate requirement R_i^{min} is considered to guarantee the QoS. Since the transmit power can affect the system performance, it could be optimized as well as the carrier allocation (i.e., $\{\alpha_{ij}, p_{ij}\}$). When we consider joint optimization and carrier and power allocation, the problem is formulated as a mixed-integer non-linear programming (MINLP) program. Finding the optimal solution of MINLP problems could require computationally complex exhaustive. In order to achieve the efficient solution, we only focus on carrier allocation with equal power allocation.

Since the formulated problem is a combinatorial one due to the binary variable α_{ij} , it could make the problem intractable for large system. Thus, we relax the carrier allocation integer constraints to take any real value in range of [0,1]. By using $\sum_{k=1}^{K} \log[U_i(\sum_{j=1}^{N} \alpha_{ij}r_{ij})]$ for (1), the objective function is log-convex, since it is a non-negative sum of log-convex functions. Since the problem is a convex one with concave feasible region, it can be solved by standard convex optimization methods. We analyze the simplified problem using the Lagrange multiplier method [17].

The Lagrangian function of the problem in (1) is expressed as,

$$L(\alpha,\lambda,\mu,\delta) = \sum_{i=1}^{K} \overline{U}_i \left(\sum_{j=1}^{N} \alpha_{ij} r_{ij} \right) + \sum_{j=1}^{N} \lambda_j \left(1 - \alpha_{ij} \right) + \sum_{i=1}^{K} \mu_i \left(\alpha_{ij} r_{ij} - R_i^{\min} \right) + \sum_i \sum_j \delta_{ij} \alpha_{ij}$$
(9)

where $\lambda = [\lambda_1, \dots, \lambda_N]$, $\mu = [\mu_1, \dots, \mu_K]$ and $\delta = [\delta_{11}, \dots, \delta_{KN}]$ are the non-negative Lagrange multiplier. Let us differentiate *L* with respect to α_{ij} and apply the Karush-Kuhn-Tucker (KKT) condition. Then, it is found that carrier *j* can be allocated to UE *i* to satisfy the following condition:

$$\mathbf{i}^* = \arg\max_{\mathbf{i}} \left[(\overline{U}_i'(r_i) + \mu_i)r_{ij} \right]$$
(10)

From (10), the optimal carrier allocation can be found when μ_i is determined for all UEs. Whilst UEs with inelastic traffic use $\mu_i > 0$, UEs of elastic traffic use $\mu_i = 0$ [18]. When the data rate is below

 R_i^{\min} , the marginal utility function of inelastic traffic is much higher than the one of elastic traffic. Based on (10), it is found that users of inelastic traffic and with better channel condition will have higher priority in carrier allocation.

Exploitation of the WiFi network

In order to find proper UEs for WiFi, we calculate the UE's access index consisting of mobility index, signal strength index [19] and traffic index. Firstly, the average received power, \bar{P}_i^w and power standard deviation, σ_i for UE *i* can be calculated as follows.

$$\bar{P}_{i}^{w} = \frac{1}{T} \sum_{t=0}^{T-1} P_{i}^{w}(t), \quad \sigma_{i}^{p} = \sqrt{\frac{1}{T} \sum_{t=0}^{T-1} \left[P_{i}^{w}(t) - \bar{P}_{i}^{w} \right]^{2}}, \quad (11)$$

where t denotes the current time slot, \bar{P}_i^w denotes the average value of received power in the latest T slots from the WiFi AP. By using the average received power and its standard deviation, we can calculate the UE's mobility level as,

$$I_i^M(t) = e^{-\sigma_i^P} \tag{12}$$

where $I_i^M(t) \in (0,1]$. $I_i^M(t)$ close to 1 indicates the a low mobility UE. The signal strength index can also be defined by received signal power and its max value as follows.

$$I_i^S(t) = \frac{P_i^W(t)}{P_{max}^W}$$
(13)

In addition, considering the traffic types of UE, the traffic index is defined as,

$$l_i^T(t) = \begin{cases} \frac{\text{\# of UEs with elastic traffic}}{\text{\# of total UEs}}, \ i = \text{elastic traffic UEs} \\ 0, \ i = \text{ inelastic traffic UEs} \end{cases}$$
(14)

The access index can be defined as a sum of weighted three indexes as follows.

$$I_i^A(t) = \alpha I_i^M(t) + \beta I_i^S(t) + \delta I_i^T(t)$$
(15)

where $\alpha + \beta + \delta = 1$ and α , β , and δ have the value range of [0,1]. They are preference factors of each index; mobility, signal strength, traffic index, respectively. For example, if the value of α is closer to 1, the allocation gives higher priority to mobility. For β closer to 1, the allocation gives higher priority to received signal strength. The higher access index of UE / indicates the UE is more beneficial to access to the WiFi AP. Then, with the UE's access index, the most proper UES for WiFi can be identified. By the definition, UEs less moving, receiving stronger signal and having elastic traffic will be identified as the proper UEs to use the WiFi network.

4.3.2 Inputs and Outputs

The Input and output variables that the proposed resource allocation algorithm will use and produce are illustrated in Table 13 and Table 14, respectively.

	Inputs					
Parameter name/ID	Description	Provided by CRRM INTERNAL block	INTERNAL interface name/ID	Provided by an EXTERNAL block	EXTERNAL interface name/ID	
UEs_C	UE's capability to access spectrum and RATs			Yes	C_5GOAM_cRRM	
UEs_T	UE requirements : 1) Traffic category & Type - e.g., (Inelastic, VoIP), 2) min data rate	-	-	Yes	C_5G-X2AP	
UEs_M	UEs' moving info : 1) location (x,y),	-		Yes	C_5G-X2AP	





	Inputs							
Parameter name/ID	Description	Provided by CRRM INTERNAL block	INTERNAL interface name/ID	Provided by an EXTERNAL block	EXTERNAL interface name/ID			
	2) direction, 3) speed							
Freq	Carriers (BW, centre freq.)	-	-	Yes	C_SM-cRRM SAP			
Rho	Channel quality, γ_{ij}	-	-	Yes	S_HMAC_cRRM SAP			
P_WiFi	The received signal	-	-	Yes	S_HMAC_cRRM SAP			
	strength from wift AP, P_i							
P_Max	Max. Tx power	-	-	-	-			
Pos_eNB	eNB's location (x,y)	-	-	-	-			
W	# of WiFi devices	-	-	Yes	S_HMAC_cRRM SAP			
P_Max_WiFi	WiFI AP's max. TX power	-	-	Yes	S_HMAC_cRRM SAP			
Pos_WiFi	WiFi AP's location (x,y)	-	-	Yes	S_HMAC_cRRM SAP			

Table 13: Input parameters of the proposed algorithm

	Outputs							
Parameter name/ID	Description	Supplied to CRRM INTERNAL block	INTERNAL interface name/ID	Supplied to an EXTERNAL block	EXTERNAL interface name/ID			
Out_A	Selected carriers for powers for UEs	-	-	Yes	M_cRRM_ConfigSAP			
Out_R	Throughput for UEs	-	-	Yes	KPI collector			
Out_U	Utility value for each UE (i.e., Qos level)	-	-	Yes	KPI collector			

Table 14: Output parameters of the proposed algorithm

4.3.2.1 Mapping of proposed algorithms to the cRRM/MAC blocks



Figure 25: Mapping the proposed algorithm to RRM functional blocks of SPEED-5G.

Figure 25 shows the mapping of the proposed algorithm. The algorithm is related to RRM functional blocks, RAT/Spectrum Selection/Aggregation, and Channel Selection, and RRM requires the



connection with the Spectrum Manager (SM) and OAM entities. While SM can provide spectrum information such as the carrier width and transmit power, OAM provides the UE's terminal capabilities including what spectrum and which RATs they are capable of handling. In addition, various data about the cell will be obtained from 5G-Cell block via C_5G-X2AP interface as shown in the figure. With the connection of KPI Collector, the algorithm can take or send the current KPI information. With the S_HMAC_cRRM SAP interface, the RRM can be provided many inputs from MAC such as channel qualities and WiFi network information delivered via XW interface from the WiFi AP. By using inputs from various entities, the proposed algorithm can decide the channels consisting of multiple carriers in licensed bands or including carriers and the link of the WiFi network. Note that the channel selection in the WiFi network is not carried out by the RRM. After deciding the channel, RRM may suggest a frame configuration to MAC via M_cRRM_Config SAP and it can send the calculated KPI to KPI collector.



Figure 26: Message sequence chart of the proposed resource allocation algorithm

The message sequence chart of the proposed algorithm is illustrated in Figure 26. The algorithm will run in a distributed way, in the dRRM of each cell. Required input including channel quality, the network loads, and the information on the WiFi networks will be collected at the RRM. Then, RRM will firstly check the resource availability (e.g., considering the WiFi network loads, whether the expected gain obtained using WiFi could be enough to increase the data rate). Then, by using the aggregation techniques, RRM will decide the channels for UEs. After the successful decision on channel selection, the output results can be sent to the MAC layer. At the MAC layer, Scheduling and Inter-RAT coordination function will be enabled and MAC will request reconfiguration to the PHY layer.

4.3.3 Simulation assumptions and parameters

By using MATLAB platform, the simulation scenario consists of one fixed eNB and a number of UEs of heterogeneous traffic types. While there is a WiFi AP, a number of WiFi devices are connected. The number of UEs and WiFi devices varies. In addition, UEs are assumed to move towards a certain



direction with the constant speed. According to [20]-[22], we select the distance-dependent path loss model: $119.6 + 37.2 \log [d](km)[dB]$ (@800 MHz), $128.1 + 37.6 \log [d](km)[dB]$ (@2GHz) and $140.7 + 37.6 \log [d](km) + 21 \log \left(\frac{2.4}{2}\right) [dB]$ (@ 2.4 GHz for the WiFi). Simulation parameters are presented in Table 15.

Parameters	Values	Parameters	Values	
ISD	500 m	# of UEs	Varying [1, 6]	
Available	8 x 5 MHz @ 800 MHz,	# of WiFi devices	Varying [1, 6]	
spectrum	10 x 2 MHz @ 2 GHz			
	1 x 20 MHz @ 2.4 GHz (WiFi)	UE' Traffic type	Mixed (Elastic, Inelastic)	
Max Tx Pwr	46 mW	UE speed	{0, 3, 10,30,60} [km/h]	
	20 mW (WiFi)			
Path Loss 800MHz: 119.6 + 37.2 log [d](km)[dB]				
2GHz:128.1 + 37.6 log [d](km)[dB]WiFi:140.7 + 37.6 log [d](km) + 21 log(2.4/2)[dB]				

Table 15: List of simulation parameters for the proposed algorithm

4.3.4 Performance Evaluation

Figure 27 shows the average utility value per UEs having different traffic. In order to compare the performance of the proposed algorithm, the algorithm is designed with $\alpha = 0$, $\beta = 1$, $\delta = 0$ in (11) as the reference. In the reference algorithm, only the signal strength of the UE is considered to access the WiFi. From Figure 27, it is shown that the proposed algorithm supports better QoS for UEs.



Figure 27: Average utility performance for different number of UEs





Figure 28: Average data rate achievable from WiFi for different UE speeds

In Figure 28, the performance on the achieved data rate from the WiFi network is illustrated for different average UE speed. The graph shows that the proposed algorithm can select the more proper UEs to access the WiFi network than the reference algorithm while UEs move with different speeds. While UEs are decided to access the WiFi network to boost the data rate, the list of proper UEs is selected in the proposed algorithm. However, the list would not be updated every second. With the static UEs (no moving), the proposed and reference algorithm have the same performance. However, when the UEs' speeds get faster, UEs selected by the reference at a certain moment could be outside the coverage of the WiFi network a bit later. Thus, although UEs of good signal strength (but moving with high speed) are selected to access the WiFi to increase their data rates, the WiFi network could not serve selected UEs. In the proposed algorithm, while the variation of the signal strength caused by the mobility is considered, mobility can be predicted. By prediction of UEs' mobility, it is highly probable that UEs less moving are selected. Thus, the proposed algorithm could generate better performance to the reference.

While the preliminary results are included in this section, sophisticated results will be included in a future deliverable.

5 RRM algorithm 4: Fuzzy MADM strategy for spectrum management in multi-RAT environments

The fifth-generation (5G) of wireless networks is being developed to meet the strict requirements of future applications (e.g., two-way gaming and the Tactile Internet). In this respect, an extensive amount of research has been devoted to developing various technologies and key enablers (e.g., ultra-densification, the design of new radio interface and use of higher frequencies) to boost the performance of future 5G radio access technologies (RATs) compared to what can be achieved using the current technologies. However, less effort has been made to ensure interworking with the existing wireless systems and standards (e.g., WLANs and LTE) and use multi-RAT operation as a complementary option to the new 5G RATs. As a matter of fact, most of the recent 5G architectures do not integrate most of the legacy RATs (e.g., GSM, WCDMA, and WLANs) or prefer to interconnect them only at the core network level because a full integration would be too costly in terms of multi-RAT measurements and interworking [23]. In such circumstances, a user-driven decision-making would be much more suitable as the user equipment (UE) could easily get a quick characterisation of all available RATs e.g., through beacons and pilot channels. The main challenge of such configuration is that the limited amount of information typically available to UEs (e.g., signal strength and SNR) is not enough to select the best RAT in a given context (e.g., QoS requirements, regulation rules, and operator policies). This clearly calls for a form of network assistance to inform UEs about all the relevant information that would be needed to perform the most efficient decision.

In this respect, it is proposed to extend the SPEED-5G architecture to support a context-aware userdriven mode of operation that may be activated for the relevant scenarios (e.g., multi-RAT environments where the various RATs do not collaborate). In particular, a connection manager (CM) is introduced on the UE side to collect the various components of the context and acts according to a policy that is remotely adjusted by the network manager. Based on this extension, a fuzzy multiple attribute decision making (MADM) implementation of the CM is developed to select the best RAT for a set of heterogeneous applications. Fuzzy logic is used to first estimate the out-of-context suitability level of each RAT to support the various application requirements. Second, an MADM component combines these estimates with the heterogeneous components of the context (e.g., user preferences and operator policies) to derive the in-context suitability levels of the considered RATs. Based on this novel metric, a spectrum management strategy combining two spectrum selection (SS) and spectrum mobility (SM) functionalities is developed to select the best RAT in a given context. As an initial use case, the proposed approach is applied to perform a context-aware offloading in dense small cell environments to support a mixture of delay-sensitive and best-effort applications.

5.1 Relation to Golden Nuggets

This piece of work contributes to the "5G Flexible Interference Mitigation and RRM" golden nugget by developing an advanced user-driven RRM strategy that efficiently selects the best available RAT and licensing regime (i.e., licensed, lightly-licensed and unlicensed) in a given context (e.g., application requirements, subscription profile and regulation rules). The proposed strategy particularly targets the use cases where the end-users are much better situated to make the most efficient decision e.g., multi-RAT environments with non-collaborating RATs and fastchanging/mobile scenarios where the selection decision has to be made fast for specific applications (e.g., delay-sensitive) or under particular circumstances (e.g., fast degradation of radio conditions). By delegating the final decision to the end-user whenever relevant, the proposed methodology provides an increased degree of flexibility in achieving a hierarchical management of ultra-dense multi-RAT networks, which falls within the scope of another golden nugget (i.e., the "5G Network Management").



5.2 Assumptions and system model

A set of *K* available RATs $({RAT_k}_{1 \le k \le K})$ is considered by various UEs to establish a set of *L* applications $({A_i}_{1 \le l \le L})$ characterised in terms a set of heterogeneous requirements $({Req_i}_{1 \le l \le L})$. Without loss of generality, *L=2* applications are considered, namely VoIP and FTP file download with stringent and loose QoS requirements, respectively. The considered environment is a hexagonal setting of LTE macro-cells overlaid by a set of buildings where several WLAN/LTE small-cells are dropped randomly (i.e., *K=2*). According to the dual-stripe layout [24], each building is modelled as two stripes of rooms with a corridor in-between, which corresponds in practice to e.g., the set of stores inside a shopping mall. The various propagation losses (i.e., indoor-indoor, outdoor-outdoor, indoor-to-outdoor and vice versa) are modelled using the hybrid building model that combines several well-known propagation loss models to consider the phenomenon of indoor/outdoor propagation in the presence of buildings [25]. As an illustrative example, Figure 29 plots the signal-to-interference-and-noise-ratio (SINR) map obtained for a co-channel configuration where a building of two stripes, of 20 rooms each, is dropped on top of a hexagonal layout of 27 LTE macrocells, with one small-cell placed in each room.



Figure 29: Illustrative example of SINR map

At the time of establishing each of the considered applications, the various contextual information that may be available about the UE (e.g., velocity, remaining power and balance) and network (e.g., operator strategy and regulation rules) should be also taken into account as they may have a strong impact on the suitability of each of the available RATs.

Therefore, the problem considered here is, whenever an application A_i needs to be established, how to:

- make the **UE** select the best RAT,
- to meet the set of application requirements,
- in the considered **context**?



5.3 Proposed context-aware user-driven framework

To achieve the target context-aware user-driven mode of operation, a functional split between the UE and network domains should be clearly made to identify the logical entities and scope of each side. In this respect, a functional architecture is first constructed and then integrated into the SPEED-5G context for an increased degree of flexibility.

5.3.1 Functional architecture

To efficiently tackle the considered problem of network selection in multi-RAT environments, the functional architecture described in Figure 30 is proposed in [26]. Specifically, a connection manager (CM) is introduced at the UE to collect the relevant components of the context from both the terminal and network sides to implement a given decision-making policy (e.g., the proposed strategy in Section 5.4). The collected contextual information is combined with a radio characterisation of each available RAT in terms of a set of short-term attributes (e.g., signal strength, SNR, and load) obtained e.g., through beacons and some medium- and long-term attributes (e.g., cost and regulation rules) stored in a policy repository together with all the policy-related parameters (e.g., used algorithm and controlling parameters). The content of the policy repository may be retrieved in practice from a local instance following a pull or push mode using e.g., the Open Mobile Alliance-Device Management (OMA-DM) protocol [27]. To offer a higher degree of flexibility to the network manager, a policy designer entity enables to build and update the policy repository content based on a set of measurement reports collected from UEs and the various network-level strategies and constraints (e.g., operator strategy and regulation rules). For instance, the policy designer may dynamically adjust some of the policy-related parameters (e.g., fuzzy logic membership functions) or the RAT attributes (e.g., cost and priority level) to optimise some network-level metrics (e.g., spectrum efficiency and energy efficiency) or implement a form of traffic steering (i.e., push UEs to use specific RATs during some periods of time). Finally, the CMs of different UEs may collaborate to further improve their individual performances, which is out of the scope of the current work.



Figure 30: Functional architecture of the proposed context-aware user-driven framework



5.3.2 Integration into the SPEED-5G architecture

This section integrates the proposed context-aware user-driven framework into the SPEED-5G architecture. This would offer a higher degree of flexibility as the proposed user-driven operation could be supported as an additional option that may be activated for the relevant use cases (e.g., non-collaborating RATs).

Figure 31 maps the functional architecture described in Figure 30 onto the relevant SPEED-5G functional modules. Specifically, the network-side functional entities (i.e., policy designer and repository) are mapped to the cRRM, with the decision-making entity (i.e., CM) being placed in the UE. Compared to the SPEED-5G default setting, this mode of operation bypasses the small cells (SCs) in terms of intelligence and delegates the final decision to the UE that acts following a policy and set of controlling parameters that are defined/adjusted on the network side. Specifically, at the networklevel, the spectrum manager and cRRM are the relevant entities as they can take advantage of their wide visibility to efficiently adjust the decision-making controlling parameters together with some of the medium- and long-term RAT attributes. At the SC level, the dRRM plays a key role in combining the adjusted parameters it receives from the cRRM with a set of short-term attributes provided by the higher-MAC. The combined data is communicated to the 5G_RRC that broadcasts it to the various UEs through a reconfiguration of some system information blocks (SIBs). At the UE level, the connection manager extracts the broadcasted information (i.e., RAT attributes and controlling parameters) from each of its physical (Phy) interfaces, and combines it with the various components of the context available locally (i.e., battery level, velocity and subscription profile) to select the best RAT. To better explain the chronological order and time-scale of each of these actions, a detailed message sequence chart will be given in Section 5.4.5.1 to describe how the algorithmic solution proposed in Section 5.4 could be in practice implemented.



Figure 31: Integration into the SPEED-5G architecture



5.4 Connection manager: Fuzzy MADM decision-making

This section implements the CM of the functional architecture described in Figure 30 to perform a context-aware exploitation of all available RATs to support each of the considered applications. To this end, the fuzzy multiple attribute decision making (MADM) strategy whose flowchart is described in Figure 32 is proposed.



Figure 32: Flowchart of the proposed fuzzy MADM strategy

Specifically, whenever an application A_i has to be established, the following three-step strategy [2] is executed:

• Rely on fuzzy logic to estimate the "out-of-context" suitability level of each RAT_k to meet the application requirements $(s_{k,l}^{oc})$ out of a set of short-term RAT attributes that is typically obtained from beacons and pilot channels (e.g., signal strength, SNR, and load).

• Based on a fuzzy MADM methodology, combine $s_{k,l}^{oc}$ with the context at hand to drive the so-

named "in-context" suitability level ($S_{k,l}^{ic}$). The context includes a set of locally available components (e.g., power level and velocity) and a set of medium- and long-term RAT attributes (e.g., cost and priority level) obtained through beacons. Some of the policy controlling parameters (e.g., MADM weights) may also be updated through beacons.

• Rank RATs in the decreasing order of in-context suitability level $s_{k,l}^{ic}$ and select the first one.

The outcome of the above strategy (i.e., selected RAT) is treated differently depending on the state of the UE. In idle mode, a request is generated to establish a connection with the selected RAT. If the request is rejected, the access procedure is iteratively performed for the next RAT in the list till the access is granted or the list of RATs gets exhausted. In connected mode, a channel reconfiguration is triggered to switch the RAT-in use if needed. To track any future change in the context, hand-overs (HOs) are triggered either on an event basis (e.g., upon QoS degradation) or periodically. In both cases, the above three-step strategy is performed again and the RAT in use is switched if needed. A continuous check for subsequent HO triggers is performed till the application ends.

For the sake of illustration, each of the above steps will be implemented to achieve the preliminary target behaviour set in Section 5.4.1. The interested reader is referred to [2] for more details about the proposed fuzzy MADM methodology and its general applicability.

5.4.1 Preliminary target behaviour

The most common scenario of a free broadband WLAN connection is considered, where indoor UEs prefer to use WLAN whenever possible and switch to the outdoor LTE macro otherwise (e.g., when indoor UEs move outside or the WLAN backhaul connection gets lost). For our particular scenario, this means that FTP file download should be always established on WLAN, while VoIP needs to maximize WLAN usage as long as its QoS requirements are met.

5.4.2 Out-of-context suitability levels

In accordance with the guidelines given in [2], two separate fuzzy logic controllers (FLCs) are designed to estimate the suitability level of each RAT to meet the VoIP QoS requirements:

- WLAN: The considered FLC together with the associated membership functions are described in Figure 33. The minimum set of input radio parameters is designed as follows:
 - 1) *SINR*: the signal-to-interference-and-noise-ratio of the AP beacon that reflects the radio and interference conditions.
 - 2) Dwell_T: the average dwell time in the AP MAC queue assumed to be advertised by its beacon. It jointly captures the channel load (i.e., due to contention) and the traffic load of the various UEs served by the AP. No 802.11e QoS support is considered initially, which means that all traffic types share the same MAC queue.
- LTE: The considered FLC and membership functions are described in Figure 34. In particular, the following radio parameters are considered:
 - *3) RSRQ*: the reference symbol received quality that captures the radio and interference conditions.
 - 4) *T_Sched*: the average time each packet waits before being scheduled. It reflects the load condition on the eNodeB and may be broadcasted in one of its SIBs. A non-QoS-aware scheduler (e.g., proportional fair (PF)) is initially assumed, which means that all packets are treated equally.





Figure 33: WLAN/VoIP fuzzy logic controller



Figure 34: LTE/VoIP fuzzy logic controller

For both FLCs, the 3²=9 required inferences rules have been designed based on a sensitivity analysis to the various combinations of the input parameters, which is omitted for the sake of brevity. This mimics the adjustment performed by the policy designer of Figure 30 based on the actual performance measurements collected from the various UEs. Finally, the defuzzification process is based on the commonly used centroid method for its accuracy [28].

Note that a possible extension to support QoS-aware bearer traffics (e.g., video over LTE (ViLTE) and voice over LTE (VoLTE)) is to make the above traffic load metrics (i.e., *Dwell_T* and *T_Sched*) separate for each application type, which is left for future consideration.

Finally, it is worth pointing out that, for the other considered application (i.e., FTP file download), there is no need to develop separate FLCs. Both WLAN and LTE are assumed to meet the loose QoS requirements as long as the corresponding UEs are associated/attached.

5.4.3 In-context suitability levels

In this section, the previously determined estimates are combined with the various components of the context to derive the in-context suitability levels. To particularly cope with the heterogeneity of the context components, an MADM methodology is developed.

The decision-maker is in our case a UE who wants to establish an application A_l and has to select among a set of alternatives (i.e., RATs). For each $k \in \{1, ..., K\}$, RAT_k is characterized in terms of the following M=4 attributes:

- ✓ $s_{k,l}^{oc}$: the out-of-context suitability to meet the set of application requirements. Recall that this is the output of the previous sub-section.
- \checkmark *cost*_{*k*} : the monetary cost of *RAT*_{*k*}.
- \checkmark power_k : the power consumption level when using RAT_k.

number.



 \checkmark range_k : an assessment of the range to reflect the appropriateness from the UE velocity perspective.

Therefore, for each application A_l , the RATs can be fully characterized in terms of a *KxM* decision matrix $\mathbf{D}_{\mathbf{l}}$ whose element $d_{k,m}^l$ denotes the performance of RAT_k in terms of the *m*-th attribute:

$$\mathbf{D}_{\mathbf{VoIP}} = \begin{pmatrix} \mathbf{QoS} & \mathbf{cost} & \mathbf{power} & \mathbf{range} \\ \mathbf{WLAN} & S_{WLAN, VoIP}^{oc} & LOW & MEDIUM & SMALL \\ \mathbf{LTE} & S_{LTE, VoIP}^{oc} & HIGH & HIGH & LARGE \end{pmatrix}$$
(16)

$$\mathbf{D}_{\mathbf{FTP}} = \begin{pmatrix} QoS & cost & power & range \\ WLAN & HIGH & LOW & MEDIUM & SMALL \\ LTE & HIGH & HIGH & HIGH & LARGE \end{pmatrix}$$
(17)

Note that, compared to LTE, WLAN is qualified as cheaper, less power-consuming and smaller in range for both applications. Recall that, for both RATs, the first attribute associated with VoIP (i.e., $S_{k,VOIP}^{oc}$) is the output of each of the FLCs designed in the last sub-section and is, therefore, a real

To adjust the relative importance of the various attributes, a vector \mathbf{w}_{I} of M weights $(\{w_{l,m}\}_{1 \le m \le M})$ is introduced for each *l*-th application:

$$\mathbf{w}_{\mathbf{VoIP}} = \mathbf{w}_{\mathbf{FTP}} = \begin{pmatrix} HIGH \\ HIGH \\ LOW \\ LOW \\ LOW \end{pmatrix}$$
(18)

Note that both $W_{l,QoS}$ and $W_{l,cost}$ are set to HIGH to achieve the target behaviour of Section 5.4.1, while $W_{l,power}$ and $W_{l,range}$ are set to LOW for the sake of simplicity (i.e., the UE power consumption and velocity are not considered initially).

Finally, the vector $\mathbf{S}_{\mathbf{I}}^{i\mathbf{c}}$ of in-context suitability levels ($\{s_{k,l}^{ic}\}_{1 \le k \le K}$) is obtained by combining the various attributes and weights as follows:

$$\mathbf{S}_{\mathbf{l}}^{\mathbf{ic}} = \begin{bmatrix} s_{\mathbf{l},l}^{ic} \\ \vdots \\ s_{k,l}^{ic} \\ \vdots \\ s_{K,l}^{ic} \end{bmatrix} = \overline{\mathbf{D}}_{\mathbf{l}} \cdot \mathbf{W}_{\mathbf{l}}$$
(19)

where $\overline{\mathbf{D}_{\mathbf{I}}}$ is the matrix of normalized attributes $\overline{d_{k,m}^{l}}$ that are calculated as $\overline{d_{k,m}^{l}} = d_{k,m}^{l} / \max_{k} (d_{k,m}^{l})$

for benefit attributes (i.e., QoS and range) and $\overline{d_{k,m}^{l}} = \min_{k} (d_{k,m}^{l}) / d_{k,m}^{l}$ for cost attributes (i.e., cost and power).

5.4.4 Decision-making

Based on the previous sub-section, the best RAT that maximises the in-context suitability level is selected for application A_i :

$$k^{*}(l) = \arg \max_{k \in \{1, \dots, K\}} (s_{k,l}^{ic})$$
(20)

To track the variability in the various attributes (e.g., radio conditions and contextual information), the CM implements the following functionalities based on the above criterion:

- **Spectrum selection (SS):** the best RAT is selected at the time of establishing each of the considered applications.
- **Spectrum mobility (SM)**: a HO to the best RAT is performed during sessions. This may be triggered on an event basis (e.g., an emergency situation due to QoS degradation) or periodically (i.e., comfort HO).

5.4.5 Inputs and Outputs

To better illustrate the scope of the proposed fuzzy MADM strategy, its sets of input and output variables are described in Table 16 and Table 17, respectively.

		Inputs			
Parameter name/ID	Description	Provided by CRRM/dRRM INTERNAL block	INTERNAL interface name/ID	Provided by an EXTERNAL block	EXTERNAL interface name/ID
SINR	Signal-to-interference- and-noise-ratio of the AP beacon (WLAN)			Yes (UE WLAN Phy).	
Dwell_T	The average dwell time in the AP MAC queue (WLAN).			Yes (UE WLAN Phy). It is set by the SC 5G_MAC and broadcasted through beacons.	
RSRQ	Reference symbol received quality (LTE)			Yes (UE LTE Phy).	
T_Sched	Average time each packet waits before being scheduled (LTE).			Yes (UE LTE Phy). It is set by the SC 5G_MAC and broadcasted through SIBs.	
Attributes of each RAT (i.e., cost, power and range)				Yes (UE Phy). It is set by the cRRM and broadcasted through beacons/pilot channels.	
Local components of the context (e.g., velocity, battery level and subscription profile)				Yes (UE internal blocks). Extracted locally by the UE dRRM (i.e., connection manager)	
MADM weights (i.e., $W_{l,QoS}$, $W_{l,cost}$,		Set locally by the UE dRRM (i.e., connection		May be altered on the network side by the cRRM.	



Inputs						
	Parameter name/ID	Description	Provided by CRRM/dRRM INTERNAL block	INTERNAL interface name/ID	Provided by an EXTERNAL block	EXTERNAL interface name/ID
	$W_{l,power}$ and $W_{l,range}$)		manager) depending on the context.			

Table 16: Input parameters of the proposed fuzzy MADM strategy

Outputs					
Parameter name/ID	Description	Supplied to CRRM INTERNAL block	INTERNAL interface name/ID	Supplied to an EXTERNAL block	EXTERNAL interface name/ID
Selected RAT	The RAT that best meets the application requirements in the context at hand			Yes: - <u>Idle mode:</u> channel request sent through the UE Phy of the selected RAT. - <u>Connected mode:</u> reconfiguration request sent to the UE RRC.	

Table 17: Output parameters of the proposed fuzzy MADM strategy

5.4.5.1 Mapping of proposed algorithms to the CRRM/MAC blocks

This section explains how the sets of input and output variables described in the previous sub-section are communicated to/from the UE CM.

The functional entities that may influence the CM operation were previously highlighted in Figure 31. To better illustrate the time-scale and chronological order of their operation, Figure 35 describes the relevant sequence of signalling messages between them. Based on a characterisation of the various bands/channels provided by the spectrum manager, the cRRM adjusts the medium- and long-term RAT attributes (i.e., cost, power and range) together with MADM weights (i.e., $w_{l,oos}$, $w_{l,cost}$, $w_{l,power}$ and $W_{l_{range}}$) on a relatively long time-scale (i.e., hundreds of milliseconds to few seconds). The adjusted parameters are provided to the SC dRRM that combines them with the set of short-term RAT attributes (i.e., Dwell T and T Sched) that are provided by the higher MAC on a shorter time-scale (i.e., few tens of milliseconds). Note that the dRRM may alter some of these attributes depending on the local constraints of the SC. The consolidated set of RAT attributes and MADM weights is then communicated to the SC 5G_RRC that updates the broadcasted SIBs of each RAT accordingly. Finally, the characterisation of each RAT is extracted by the associated UE Phy interface and fed into the CM that combines them with the local components of the context to determine the best RAT. If the UE is in idle mode, a channel request is sent to establish a connection through the Phy interface of the selected RAT. Otherwise (i.e., UE is in connected mode), a channel reconfiguration request is sent to the UE RRC to switch the in-use RAT whenever the serving and target RATs are different.







To get insight into the effectiveness of the proposed methodology in supporting traffic offloading, the L=2 applications considered in Section 5.2 are modelled as follows:

- VoIP: The VoIP traffic model is based on G.729A. It generates packets of 60 Bytes (i.e., payload plus IP header overhead) at an inter-arrival time of 20ms, which corresponds to a data rate of 24 Kbps. The associated set of QoS requirements Req_1 is composed of a maximum end-to-end delay of D_{max} =150 ms and frame loss ratio of L_{max} =5%. In this respect, the VoIP receiver accepts only in-sequence frames whose end-to-end delay does not exceed D_{max} . Any other frame is dropped with no subsequent retransmission.
- FTP file download: an ON/OFF model is used to model file download sessions (i.e., ON periods) and the inactivity intervals in-between (i.e., OFF periods). Both ON/OFF periods are exponentially distributed with rate λ . Whenever a file download session is established, it uses the whole capacity of the in-use radio link (i.e., WLAN or LTE) with loose QoS requirements $(Req_2 = \emptyset)$.

5.4.7 Simulation assumptions and parameters

The simulation parameters for the performance evaluation of the proposed fuzzy MADM strategy are summarized in Table 18.

Parameters	Values		
Layout	Dual-stripe (see Section 5.2)		
Number of considered rooms	1 (i.e., small cells and UEs dropped inside the same room)		
K (number of RATs)	2 (i.e., LTE and WLAN)		
Number of LTE macrocells	27		
LTE mode	FDD		
Carrier frequency of LTE macrocells	2.120 GHz		
Number of WLAN APs	1		
WLAN standard	802.11n		
WLAN channel	36 (i.e., 5.180 GHz)		
Full list of physical parameters	See outdoor scenario of Annex A in [29]		
L (number of applications)	2 (i.e., file download and VoIP)		
Traffic model	See Section 5.4.6		
Number of UEs per application	1		
Direction of traffic	downlink-only		
Simulation time	10s		

Table 18: Simulation assumptions and parameters for evaluating the fuzzy MADM strategy

5.4.8 Benchmarking

To assess the influence of the different components of the proposed framework, the following variants will be compared:

speed5



- *Initial+Emergency*: The fuzzy MADM selection scheme proposed in Section 5.4.4 is applied both initially and during sessions. Only emergency HOs are triggered (i.e., when QoS requirements are not met).
- Initial+Emergency+Comfort: In addition to the triggers of Initial+Emergency, periodic comfort HOs are triggered each Δt . If a better RAT_{k^*} is identified (i.e., $s_{k^*l}^{ic} > s_{serving,l}^{ic}$), the UE is reconfigured to use it. To avoid ping-pong effects, comfort HOs are blocked whenever any of the QoS requirements is not met.
- *WLAN if coverage*: a benchmarking scheme that always selects WLAN if the corresponding UE is associated.

5.4.9 Performance evaluation

This section evaluates the performance of the proposed fuzzy MADM approach in achieving a context-aware offloading in the considered environment. In this respect, a set of system-level simulations have been performed using the NS-3 simulator.

Figure 36 shows the time evolution of the end-to-end delay of the VoIP application for all variants introduced in Section 5.4.8 with the threshold D_{max} shown in dashed lines. For better analysis of the obtained results, constant ON/OFF durations for FTP sessions are initially considered (i.e., $1/\lambda=2s$) with comfort HOs triggered each $\Delta t=0.5s$.



Figure 36: Evolution of the end-to-end delay of VoIP





(a) WLAN usage fraction of VoIP (b) Frame loss ratio of VoIP

Figure 37: Analysis of the effectiveness in achieving the target behaviour

The results show that *Initial+Emergency* introduces a significant improvement in terms of the end-toend delay compared to the baseline *WLAN if coverage*. After VoIP is initially established, the WLAN link gets quickly saturated when the first FTP session starts (i.e., around t=3s), which strongly degrades the end-to-end delay for *WLAN if coverage*. In turn, when *Initial+Emergency* is used, an emergency HO to the macro LTE is triggered shortly after the VoIP receiver starts to receive some frames whose end-to-end delay exceeds the threshold $D_{max}=150ms$. Given that all subsequent FTP sessions are established on WLAN, *Initial+Emergency* will not experience any further degradation and no subsequent HO will be triggered. On the contrary, shortly after the end of the FTP session, *Initial+Emergency+Comfort* triggers a comfort HO back to WLAN (i.e., around t=5.4s), which makes it subject to a future degradation when the next FTP session starts before a new emergency HO can be triggered (e.g., around t=7.3 s).

To better illustrate the effectiveness in achieving the target behaviour, Figure 37(a) plots the WLAN usage fraction achieved by each of the considered schemes for different average ON/OFF durations. Figure 37(b) shows the corresponding average frame loss ratio (FLR) with the threshold L_{max} shown in dashed lines.

The analysis of the observed behaviour reveals that *Initial+Emergency* introduces a significant improvement in terms of FLR compared to *WLAN if coverage*, which confirms the previous observation in Figure 36. When comparing *Initial+Emergency* and *Initial+Emergency+Comfort*, it can be seen that the additional comfort HOs significantly increase the fraction of using WLAN with gains up to 160% (Figure 37(a)). When long ON/OFF durations are used (i.e., $1/\lambda>2s$), few comfort HOs are executed, which keeps the number of lost frames relatively low, and maintains an acceptable FLR (Figure 37(b)). As a result, comfort HOs are never blocked, which justifies the higher fraction of using WLAN achieved by *Initial+Emergency+Comfort* in Figure 37(a). When shorter ON/OFF durations are used, more comfort HOs are initially executed with the associated increase in the number of lost frames. Interestingly, the FLR does not further degrade but stabilises around the requirement L_{max} as can be observed in Figure 37(b). This is because as soon as FLR degrades, comfort HOs back to WLAN are blocked, which helps to increase the WLAN usage fraction only to the extent that does not hurt the FLR requirement.

In summary, *Initial+Emergency* avoids selecting WLAN when QoS requirements are not met, which significantly outperforms the traditional offloading in terms of FLR. Executing comfort HOs on top of it further increases WLAN usage to the maximum extent that does not hurt any of the QoS requirements, which efficiently achieves the target behaviour.

6 RRM algorithm 5: Co-primary spectrum sharing in uplink SC-FDMA networks

6.1 Relation to Golden Nugget

This co-primary spectrum sharing algorithm is related to the "High-level design of RRM functions and their interfaces to one another" Speed-5G Golden Nugget. The proposed algorithm allows mobile network operators (MNOs) who employ infrastructure sharing to efficiently share the available spectrum resources, taking advantage of information coming from the Physical and MAC layers, in order to avoid inter-operator interference and achieve improved Quality of Service (QoS) for real-time applications.

6.2 Assumptions and system model

We consider a common pool of shared spectrum, for the case of MNOs operating in the uplink direction of LTE and employing infrastructure sharing. The objective of the spectrum sharing process is to assure exclusive access to the shared spectrum in order to avoid inter-operator interference. This scheme can be applied to both LTE macro and small cells. The system model consists of a single LTE cell and a number of User Equipment (UE) devices, each one associated with an MNO, randomly deployed in the cell coverage area. For the remainder of this section, the terms user and UE are used interchangeably. Each user has an active real-time video connection on the uplink and the eNodeB / Access Point is responsible for allocating the available resources in a fair, QoS and energy efficient manner, employing the proposed scheme.

6.3 Algorithm Description

6.3.1 Algorithm flowchart

Since, as discussed in the theoretical formulation of the co-primary spectrum sharing problem in D4.1 [2], the optimal allocation of uplink scheduling blocks in a localized SC-FDMA LTE system is an NP-hard problem, in this section we introduce a suboptimal algorithm that takes into consideration the users' buffer status and real-time delay constraints, as well as the operator priorities and the constraints of a realistic LTE system in order to perform uplink resource allocation in a QoS and energy efficient manner.

On the uplink direction of an LTE network, resource allocation is performed on a per subframe basis. In order to perform resource allocation in a fair, QoS and energy efficient manner and evaluate the utility of each scheduling block of the shared spectrum pool to the users of each MNO, we introduce metric $m_{i,j,k}^{UL}(t)$ of user $i, i \in UE$, who is a subscriber of MNO k, for scheduling block $j, j \in \{1, 2, ..., N_{SB}\}$, where N_{SB} is the number of scheduling blocks per subframe of the shared spectrum, as follows:

$$m_{i,j,k}^{UL}(t) = p_k \frac{d_{i,k}^{UL}(t)}{d_{th,i,k}} \exp\left(\frac{\overline{D}_{i,k}^{UL}(t)}{\overline{R}_{i,k}^{UL}(t)}\right) \frac{r_{i,j,k}^{UL}(t)}{P_{1,i,k} \cdot N_{RB,SB}}.$$
(21)

 p_k is the shared spectrum access priority of MNO k. This can be either a static parameter, or a dynamic value that depends on the performance of spectrum allocations in the past. $d_{i,k}^{UL}(t)$ is the time passed since the last uplink grant was allocated to user i or since a Scheduling Request (SR) has been received from this user and $d_{th,i,k}$ is the delay threshold, beyond which a packet of a real-time application is no longer considered usable and is discarded by the user's buffer. Since the eNodeB/AP does not have accurate information on the exact waiting time of the pending packets of each user, $d_{i,k}^{UL}(t)$ is used in order to allow a worst-case estimation of the packet delay, i.e., the case of a

new packet entering the user's uplink buffer just after an uplink grant was allocated to the user or a SR was sent. Therefore, with the use of $d_{i,k}^{UL}(t)$, the prioritization of users who have waited for a higher amount of time since their last uplink grant or the latest SR, and are in higher risk of packet expiration, is achieved. $\overline{D}_{i,k}^{UL}(t)$ and $\overline{R}_{i,k}^{UL}(t)$ are the average delay and data rate, respectively, experienced by user *i* in the past, and are calculated using a weighted moving average formula as follows:

$$\overline{D}_{i,k}^{UL}(t) = \beta d_{i,k}^{UL}(t) + (1 - \beta) \overline{D}_{i,k}^{UL}(t - 1),$$
(22)

$$\overline{R}_{i,k}^{UL}(t) = \beta r_{i,k}^{UL}(t) + (1 - \beta) \overline{R}_{i,k}^{UL}(t - 1),$$
(23)

where $r_{i,k}^{UL}(t)$ is the instantaneous uplink data rate of user i and $0 \le \beta \le 1$. The incorporation of $\overline{D}_{i,k}^{UL}(t)$ and $\overline{R}_{i,k}^{UL}(t)$ in $m_{i,j,k}^{UL}(t)$ allows the prioritization of users that were served with high average delay and low average data rate in the past, thus increasing the fairness of the proposed solution. $P_{1,i,k}$ is the minimum uplink power per resource block of user i, which, based on the LTE uplink power control specification, is defined as follows:

$$P_{1,i,k} = \min \left\{ P_{CMAX,C}, \ P_{0,PUSCH} + \alpha P L_{i,k} + 10 \ \log_{10}(N_{RB}^{UL}) \right\} - 10 \ \log_{10}(N_{RB}^{UL}).$$
(24)

 $P_{1,i,k}$ is calculated based on the assumption that all the resource blocks of an uplink slot are allocated to user *i*. Of course, the actual uplink power per resource block will almost always be higher for the specific user, and will depend on the actual number of its allocated resource blocks, which, in principle, will be less than N_{RB}^{UL} . $P_{CMAX,C}$ is the configured UE transmit power, $P_{0,PUSCH}$ is the target received power per resource block, while $PL_{i,k}$ is the user downlink path loss estimate calculated in the UE and α , $0 \le \alpha \le 1$, is a parameter for path loss compensation whose value is provided by the higher layers [30]. $r_{i,j,k}^{UL}(t)$ is the data rate achieved by user *i* on scheduling block *j* and is defined as follows:

$$r_{i,j,k}^{UL}(t) = \frac{\left(L_{SB}^{UL}\log_2 M_{i,j,k}\right)}{T_{sf}}$$
(25)

where L_{SB}^{UL} is the number of data carrying resource elements per uplink scheduling block, which depends on the number of reference signals transmitted in a subframe, $M_{i,j,k}$ is the Modulation and Coding Scheme (MCS) of user *i* on scheduling block *j* and T_{sf} is the subframe length. In a generic SC-FDMA system that allows the selection of different MCS per scheduling block based on the perceived channel conditions, the value of $r_{i,j,k}^{UL}(t)$ is different for every scheduling block. However, since according to the LTE system specifications all scheduling blocks allocated to the same user have the same MCS, the value of $r_{i,j,k}^{UL}(t)$ and, consequently, the value of $m_{i,j,k}^{UL}(t)$ will be the same across all scheduling blocks.

As a first step, the set of active users UE is sorted in descending order of $m_{i,k}^{UL}(t)$. This is a metric that aims to provide higher resource allocation priority to users with increased operator priority, waiting time with respect to the delay threshold, high average delay and low average data rate of their allocations in the past, as well as low uplink power transmission requirements and high expected data rate per scheduling block. To this end, $m_{i,k}^{UL}(t)$ is defined as follows:

$$m_{i,k}^{UL}(t) = p_k \frac{d_{i,k}^{UL}(t)}{d_{th,i,k}} \exp\left(\frac{\overline{D}_{i,k}^{UL}(t)}{\overline{R}_{i,k}^{UL}(t)}\right) \frac{1}{P_{1,i,k} \cdot N_{RB,SB}} E[r_{i,j,k}^{UL}(t)].$$
(26)

The operation of the proposed co-primary spectrum sharing algorithm in each subframe of length equal to T_{sf} is formally described in Table 19 and depicted in the flowcharts of Figure 38 and Figure 39. The algorithm iterates until either all the scheduling blocks of the subframe are allocated, i.e., the set Φ of available scheduling blocks is empty, or all users have received enough resources to accommodate their uplink transmission needs, i.e., the set UE of active users is empty. Therefore, for each user $i \in UE$, in descending order of $m_{i,k}^{UL}(t)$, the proposed algorithm performs the following steps:



- 1) Firstly, the user's need for an uplink transmission grant is assessed. This is based on whether a SR is received by the user, i.e., $SR_{i,k}(t) = 1$, or the value of the latest Buffer Status Report (BSR) verifies that the user buffer has uplink data waiting to be transmitted, i.e., $BSR_{i,k}(t) > 0$. If there is no need to allocate uplink resources in this subframe, the user is removed from UE and the algorithm proceeds to the next user.
- 2) If either $SR_{i,k}(t) = 1$ or $BSR_{i,k}(t) > 0$ the resource allocation algorithm determines the set $K_{i,k}$, which consists of the available scheduling blocks for which the user maximizes the value of $m_{i,j,k}^{UL}(t)$, i.e., $K_{i,k} = \left\{j' \in \Phi: i = \arg \max_{i' \in UE} \left(m_{i',j',k'}^{UL}(t)\right)\right\}$. It has to be noted that the scheduling blocks that comprise $K_{i,k}$ are not necessarily contiguous.
- 3) If $K_{i,k}$ is nonempty, the scheduling block j^* with the highest SNR $\gamma_{i,j,k}$ is determined, i.e., $j^* = \arg \max_{j \in K_{i,k}} (\gamma_{i,j,k})$ and, if its BER, i.e., $BER_{i,j^*,k}$, is lower than the threshold τ , it is the first scheduling block to be included in set $G_{i,k}$, i.e., the set of all scheduling blocks allocated to user i in this subframe.



Figure 38: Flowchart of the proposed uplink co-primary spectrum sharing algorithm in each Time Transmission Interval (TTI).

4) The set $G_{i,k}$, which contains scheduling block j^* , as well as the maximum number of contiguous scheduling blocks neighboring j^* that can be allocated to user i is calculated. This depends on the user's buffer status, the availability of scheduling blocks that are neighbors to j^* , as well as on

Speed 55

the value of $m_{i,j,k}^{UL}(t)$. Therefore, a scheduling block j is included in set G_i , if i) it is not already allocated to another user, i.e., $j \in \Phi$, ii) it maximizes the value of $m_{i,j,k}^{UL}(t)$, i.e., $i = \arg \max_{i' \in UE} \left(m_{i',j,k}^{UL}(t) \right)$, iii) it is a neighbor to another scheduling block that is already included in G_i , therefore not violating the scheduling block contiguity constraint, i.e., $\exists j' \in G_{i,k}: |j - j'| = 1$, iv) its BER is lower than the threshold τ , and v) the number of scheduling blocks already included in G_i is not enough to accommodate all the traffic in the user's buffer, which is depicted as $L_{BSR} \left(BSR_{i,k}(t) \right)$. The number of bytes that can be accommodated by scheduling block j depends on the user's MCS and is depicted as $L_{SB}(j)$. In order to determine the scheduling blocks that comprise G_i , the proposed algorithm uses j^* as a starting point and attempts to expand the allocation towards both directions, i.e., scheduling block that does not qualify one or more of the above five criteria for inclusion in G_i is met. The detailed steps of this process are described in Table 19 and the flowchart of Figure 39.



Figure 39: Flowchart of the calculation of $G_{i,k}$



Algorithm: Uplink Co-Primary Spectrum Sharing

Sort UE in descending order of $m_{i,k}^{UL}(t)$, $\forall i \in UE$ Calculate $m_{i,j,k}^{UL}(t), \forall i \in UE, j \in \{1, 2, ..., N_{SB}\}, \forall k \in MNO$ for $i \in UE$ do if $\Phi \neq \emptyset$ then $G_{i,k} \leftarrow \emptyset$ if $BSR_{i,k}(t) > 0$ or $SR_{i,k}(t) = 1$ then $K_{i,k} = \left\{ j' \in \Phi: i = \arg \max_{i' \in UE} \left(m_{i',j',k'}^{UL}(t) \right) \right\}$ if $K_{ik} \neq \emptyset$ then $j^* \leftarrow \arg\max_{j \in K_{i,k}} (\gamma_{i,j,k}), BER_{i,j^*,k} < \tau$ $G_{i,k} \leftarrow G_{i,k} \cup \{j^*\}$ $L_{i,k} \leftarrow L_{BSR} \left(BSR_{i,k}(t) \right) - L_{SB}(j^*)$ $j \leftarrow j^* + 1$, end $\leftarrow 0$ while $j \in \Phi$ and $L_{i,k} > 0$ and end = 0 do if $i = \arg \max_{i' \in UE} \left(m^{UL}_{i',j,k}(t) \right)$ and $BER_{i,j,k} < \tau$ then $G_{i,k} \leftarrow G_{i,k} \cup \{j\}$ $L_{i,k} \leftarrow L_{i,k} - L_{SB}(j)$ $i \leftarrow i + 1$ Else end $\leftarrow 1$ End end $i \leftarrow i^* - 1$, end $\leftarrow 0$ while $j \in \Phi$ and $L_{i,k} > 0$ and end = 0 do if $i = \arg \max_{i' \in UE} \left(m_{i',j,k}^{UL}(t) \right)$ and $BER_{i,j,k} < \tau$ then $G_{i,k} \leftarrow G_{i,k} \cup \{j\}$ $L_{i,k} \leftarrow L_{i,k} - L_{SB}(j)$ $j \leftarrow j - 1$ else end $\leftarrow 1$ end end end end $UE \leftarrow UE \setminus \{i\}$ $\Phi \leftarrow \Phi \setminus \{G_i\}$ end end

Table 19: Uplink so-primary spectrum sharing algorithm pseudo code

6.3.2 Inputs & Outputs

The proposed co-primary spectrum sharing algorithm operates exclusively in the dRRM. Since it operates per TTI and there is no cRRM involvement considered, we assume that the RRM is co-located with the Higher MAC and that any communication between them takes place via internal interfaces.

The inputs and outputs of the proposed algorithm are provided in Table 20 and Table 21, respectively.


Parameter name/ID	Description	source CRRM INTERNAL block	INTERNAL interface name/ID	source EXTERNAL block	EXTERNAL interface name/ID
p_k	Priority of MNO k				
$d_{i,k}^{UL}(t)$	Estimated uplink queuing				
	delay of user <i>i</i> , operator				
	<i>k</i> (s)				
$d_{th,i,k}$	Queuing delay threshold				
	of user i , operator k (s)				
$\overline{D}_{i,k}^{UL}(t)$	Average uplink delay of				
	user i , operator k (s)				
$r_{i,i,k}^{UL}(t)$	Instantaneous uplink				
	data rate of user <i>i</i> on				
	scheduling block j,				
	operator k (b/s)				
$\bar{R}_{i,k}^{UL}(t)$	Average uplink data rate				
-,	of user i , operator k (b/s)				
P _{CMAX,C}	Configured UE transmit				
	power				
P _{0,PUSCH}	Target received power				
	per resource block				
$PL_{i,k}$	User downlink path loss				
	estimate				
N_{RB}^{UL}	Total number of resource			yes	C_HMAC_LMAC
	blocks per slot				SAP
N _{RB.SB}	Number of resource			yes	C_HMAC_LMAC
, -	blocks per scheduling				SAP
	block				
Yi.i.k	Signal-to-Noise Ratio of				
,,,,,-	user <i>i</i> on scheduling				
	block <i>j</i> , operator <i>k</i>				
BSR _{i.k}	Buffer Status Report of			yes	C_HMAC_LMAC
-,	user <i>i</i> , operator <i>k</i>				SAP
$SR_{i,k}(t)$	Scheduling Request of			yes	C_HMAC_LMAC
.,	user <i>i</i> , operator <i>k</i>				SAP

Table 20: Input parameters of the uplink co-primary resource allocation algorithm

		Outputs	5		
Parameter name/ID	Description	Supplied to CRRM INTERNAL block	INTERNAL interface name/ID	Supplied to an EXTERNAL block	EXTERNAL interface name/ID
$Sch = \{SB_j, UE_i, MNO_k\}$	The outcome of the resource allocation. Each entry of the array represents a scheduling block, the user and the MNO it is allocated to.			yes	C_HMAC_LMAC_SAP

Table 21 Output parameters of the uplink co-primary resource allocation algorithm



The Message Sequence Chart of the proposed algorithm is shown in Figure 40. The algorithm, which operates in the dRRM, first collects the necessary inputs from the L_MAC and H_MAC blocks, see Table 20. Based on this information, it allocates the available resources to the different users of each MNO. The outcome of this operation is provided to the L_MAC.



Figure 40: Message Sequence Chart of the uplink co-primary spectrum sharing algorithm

6.3.3 Performance measures/KPIs

The performance of the proposed uplink co-primary spectrum sharing algorithm is measured in terms of packet timeout rate, goodput, fairness, and average delay.

The packet timeout rate is defined as the percentage of packets that are discarded due to excessive delay in the unit of time.

The goodput is defined as the throughput at the application layer, i.e., the rate of useful bits that reach the application layer in the unit of time.

Fairness is evaluated using the Jain Index of Fairness, i.e., $FI = \frac{(\sum_{i \in UE} Th_i(t))^2}{|UE| \cdot \sum_{i \in UE} Th_i^2(t)}$, where $Th_i(t)$ is the throughput of user *i* [31].

6.3.4 Simulation assumptions and parameters

The simulation parameters for the performance evaluation of the proposed algorithm are summarized in Table 22.

Parameters	Values
Physical layer parameters	Channel Bandwidth: 10MHz,
	Subframe length (T_{sf}) : 1ms,
	Number of resource blocks (N_{RB}^{UL}): 50
Resource block format	Number of subcarriers per resource block (N_{SC}^{RB}) : 12, Number of symbols per resource block (N_{symb}^{UL}) : 7, Subcarrier spacing: 15kHz
Reference Signal transmissions	2 Reference Signal transmissions per subframe
TDD configuration	Configuration 1, DL:UL 3:2



Parameters	Values
Modulation and Coding Schemes	QPSK ½, 16-QAM ½ and 64-QAM ¾
Path loss model [32]	128.1+37.6 $\log_{10} d$, d : distance from the eNodeB (km)
Transmitter antenna gain [33]	OdBi
Receiver antenna gain [33]	18dBi
Cable loss [33]	OdB
Receiver Noise Floor [33]	-116.4dBm
Interference Margin [33]	1dB
Control Channel Overhead [33]	OdB
Shadowing	Log normal, σ =8dB
Fading	Rayleigh
Maximum UE transmission power	23dBm
Target received power $(P_{0,PUSCH})$	-57dBm
Uplink path loss compensation factor (α)	0.7
Maximum tolerable delay ($d_{th,i}$)	20ms
RLC mode	Unacknowledged mode (UM)
Traffic model [34]	H264 video traffic QCIF 176×144
Protocol header sizes	RTP/UDP/IP with ROCH Compression: 3 bytes, PDCP: 2 bytes, RLC: 3 bytes, MAC: 2 bytes, CRC: 3 bytes
Moving average calculation factor (eta)	0.2
Maximum distance from the eNodeB	330m
Simulation time	67s

Table 22: Simulation Parameters of the Uplink Co-Primary Spectrum Sharing Algorithm

In order to achieve statistical accuracy, 100 simulation runs were executed.

6.3.5 Performance Evaluation

Figure 41 depicts the average packet timeout rate versus an increasing number of users and the shared spectrum access priority of MNO_1 , which is referred to as p_1 . The priority of MNO_2 , is defined as follows: $p_2 = 1 - p_1$. The total number of users N is in the range of [15,35], while the number of users of each MNO is defined as follows: $N_1 = \left\lfloor \frac{N}{2} \right\rfloor$ and $N_2 = N - N_1$. The packet timeout rate is defined as the number of packets that expire in the unit of time, since in real-time applications excessive scheduling delay leads to discarding of expired packets. As it can be seen, for both MNOs the packet timeout rate follows an increasing course with the number of users, since an increasing number of users results in higher congestion, which leads to longer queuing time and, eventually to more packet expirations. The effect of the shared spectrum priority, p_k , is also shown in this figure. As it can be seen, the mean packet timeout rate of MNO_1 follows a declining course with the increase of p_1 , while the packet timeout rate of MNO_2 follows the opposite course, since $p_2 = 1 - p_1$. This is a result of the fact that the proposed algorithm takes into consideration the shared spectrum access priority of each MNO during the resource allocation.





Figure 41: Average packet timeout rate of (a) MNO1 and (b) MNO2 versus the number of users and the shared spectrum access priority.

Figure 42 depicts the average packet delay versus the number of users and the shared spectrum access priority. As it can be seen, for both MNOs, the average packet delay follows an increasing course with the number of users, as a result of the increased congestion and the need for longer queuing times. Moreover, it can be seen that the average delay follows a declining course with the increase of the shared spectrum priority of the MNO, due to the fact that the proposed algorithm prioritizes users based on the priority of the MNO they are associated with.

Figure 43 depicts the average goodput of both MNOs. The goodput is defined as the throughput at the application layer, i.e., the rate of useful bits that reach the application layer in the unit of time. As it can be seen, in both cases the goodput follows a declining course with the increase of the number of users, as a result of the increasing congestion, which leads to excessive packet delays and timeouts. Moreover, it is shown that the average goodput increases with the increase of the MNO spectrum access priority.





Figure 42: Average delay of (a) MNO1 and (b) MNO2 versus the number of users and the shared spectrum access priority.

Figure 44 depicts the fairness of both operators with respect to an increasing number of users of both operators and the shared spectrum access priority. Fairness is evaluated using the Jain Index of Fairness, see section 6.3.3 above. As it can be seen, fairness of the system that employs the proposed algorithms is considerably high, as a result of the fact that the proposed algorithm takes into consideration the average packet delay $\overline{D}_{i,k}^{UL}(t)$ and the average data rate $\overline{R}_{i,k}^{UL}(t)$ in the user prioritization, therefore favoring users that have experienced high average delay and low average data rate in past allocations.









Figure 43: Goodput of (a) MNO1 and (b) MNO2 versus the number of users and the shared spectrum access priority.



Figure 44: Fairness (Jain index) versus the total number of users and the MNO priority.



7 RRM algorithm 6: Dynamic resource allocation algorithms for coexistence of LTE-U and WiFi

Due to the exponential increase in mobile data, efficient spectrum utilisation is the most essential resource for the mobile operators now a days. Therefore, mobile operators have been offloading more and more data traffic from their overloaded networks to a large number of WiFi hotspots over the past year. The approach of utilizing both licensed band and unlicensed band has helped mobile operators to narrow the gap between the limited capacity of cellular network and the fast growing demand of mobile broadband.

However, with a significant amount of unlicensed spectrum globally available in the 5 GHz band, the mobile operators and vendors are looking to use unlicensed spectrum to augment the capacity of licensed frequency carriers. This new way to access 5 GHz is formally known as LTE-Unlicensed (LTE-U). For operators, LTE-U means synchronised integrated network management, same authentication procedures, more efficient resource utilisation and thus lower operational costs. For wireless users, LTE-U means enhanced user experience, i.e. higher data rates, seamless service continuity between licensed and unlicensed bands, ubiquitous mobility and improved reliability. Nevertheless, it is observed that the coexistence of LTE-U and WiFi in the same frequency bands causes a meaningful degradation on the system performance. Currently, WiFi systems adopt a contention-based MAC protocol with random back-off mechanism. If left unrestrained, unlicensed LTE transmissions can actively and aggressively occupy the channel (i.e. 5 GHz) and make the medium busy most of the time. This will not only degrade the WiFi devices throughput but also overall throughput of the system [35].

Therefore we are proposing an algorithm which maximize network throughput in multi-operator scenario for 5G mobile systems by jointly considering licensed & unlicensed band, user association and power allocation subject to minimum rate guaranty and co-channel interference threshold.

7.1 Relation to GN

We are considering the ultra-dense scenario operating in multi-RAT, multiband and multi-operator environment.



Figure 45: Dynamic resource allocation algorithms scenario.



All the operations of considered scenario is managed centrally, which is one of the golden nuggets that have been proposed by SPEED-5G in the 5G-PPP association. In SPEED-5G centralized management is used as a baseline as explained in section 2 which can be expanded with distributed management by moving management decisions closer to the node. In line with SPEED-5G RRM framework, the proposed algorithm will run fully in centralized manner and make decision related to RAT, spectrum, and channel selection for each operator.

7.2 Assumptions and system model

7.2.1 System Model

We consider multi-operator LTE-U HetNet, where two operators (A&B) have non-collocated macrocell, smallcell, and WiFi Access point (AP) as shown in Figure 45. We consider non-ideal backhaul between smallcell and macrocell. We also assume that UE in an operator will be served by one base station (BS) among macro eNB on licensed band, the small cell on licensed band/unlicensed band (5GHz) and WiFi AP on the unlicensed 5GHz band for downlink transmissions. Let $x_i^{(k,o)}$ be a binary indicator to show UE k is connected to which operator, operator A or operator B, i.e $o \in O$ and $O = \{A, B\}$. It also indicates UE k is connected to whom among the macro eNB on the licensed band (m_i) , the small cell on the licensed band (s_i) , the small cell on the unlicensed band (s_u) and WiFi on the unlicensed band (w_u) , i.e $i \in l$ and $l = \{m_l, s_l, s_u, w_u\}$. This indicator can be described as

$$x_{i}^{(k,o)} = \begin{cases} 1, if \ UE \ k \ is \ connected \ to \ BS \ i \ in \ operator \ o \\ 0, otherwise \end{cases}$$
(27)

Initially, UEs connected to the small cell will be served on licensed band when the traffic load is increased, LTE-U interface is turned-on in a small cell and UEs in the small cell are served with 5 GHz unlicensed band, same as in WiFi AP. For simplicity, we assume that while sharing licensed band between small cell and macro cell, different orthogonal channels are used to avoid strong interference. Let K be the number of UEs that want to communicate with each other. The channel gain between k^{th} UE and BS i is h_i^k , G_o be the antenna gain and $\zeta 10^{\zeta/10}$ be the lognormal shadowing, where ζ is the zero mean gaussian random variable with standard deviation σ , and the

channel gain h_i^k is modeled as $h_i^k = \overline{h_i^k} \zeta G_o \left(\frac{d_o}{d}\right)^{\alpha}$ where d is the distance between transmitter and

receiver, d_o denotes antenna far field reference distance, α is path loss exponent, \overline{h}_i^k is Rayleigh random variable.

7.2.2 Assumption

We assume that LTE operates offload data traffic in licensed band to unlicensed band by deploying small cell, which operated in both LTE and LTE-U. The coverage of small cells is approximately equal to WiFi access points as shown in Figure 45. The LTE unlicensed band transmission, located at 5.8 GHz with bandwidth 20 MHz, is assisted by licensed band access and transmit only downlink non-GBR user data. WiFi shared the same 20 MHz band with both downlink and uplink using TDD. All the equipment are under the coverage of each other, which means no hidden terminal problem. Small cell, WiFi, and users are always backlogged, which means they always have data ready to transmit. We assume no retry limit for WiFi and users (no ARQ), the constant bit rate for LTE-U small cell, WiFi, and users.

Moreover, we assume that LTE and LTE-U interact with each other through carrier aggregation. The primary component carrier (PCC), which is in charge of RRC connection establishment, while the



secondary component carrier (SCC), which is only for downlink non-GBR user data transmission, is located in an unlicensed band.

Symbols	Definition
k	Number of Users (UE)
t _u	Portion of time unlicensed band is used by small cell UE
t_l	Portion of time licensed band is used by small cell UE
$P_{s_{u}}$	Max. the power of small cell on unlicensed band
P _{wu}	Max. the power of WiFi A.P on unlicensed band
P_{s_l}	Max. the power of small cell on licensed band
P_{m_l}	Max. the power of macro cell on licensed band
$p_{s_u}^k$	the power of user K connected to small cell on unlicensed band
$p_{w_u}^k$	the power of user K connected to WiFi A.P. on unlicensed band
$p_{s_l}^k$	the power of user K connected to small cell on licensed band
$p_{m_l}^k$	the power of user K connected to a macro cell on licensed band
R_{s_u}	Sum-rate of UEs on small cell on unlicensed band
R _{sl}	Sum-rate of UEs on small cell on licensed band
$R_{_{W_{_{u}}}}$	Sum-rate of UEs on WiFi AP on unlicensed band
R_{m_l}	Sum-rate of UEs on macro cell on licensed band
R^k	Min required rate
$g_{s_u}^k$	Channel gain between UE and small cell on unlicensed band
$g_{s_l}^k$	Channel gain between UE and small cell licensed band
h_m^k	Channel gain between UE and macro cell on licensed band
e_w^k	Channel gain between UE and WiFi A.P on unlicensed band
$f_{{\scriptscriptstyle W},{\scriptscriptstyle s}_u}^{j,k}$	Channel gain between UE k on small cell $S_u^{}$ and UE j on WiFi A.P
$q_{s_u s_u}^{j,k}$	Channel gain between UE k on small cell $S_u^{}$ and UE j on another small cell $S_u^{}$
$q^{j,k}_{{}^{w_uw_u}}$	Channel gain between UE k on WiFi $ \mathcal{W}_{\! u} $ and UE j on another WiFi $ \mathcal{W}_{\! u} $
$B_{_{W}}$	Total bandwidth on unlicensed band
B_l	Total bandwidth on licensed band
N_o	Additive white Gaussian noise
G_{o}	Antenna gain
ς	Zero-mean Gaussian random variables for shadowing
<i>C</i> 1– <i>C</i> 11	Constraint 1 to Constraint 11

Table 23: Notations

7.2.3 Achievable data rate in unlicensed spectrum

The basic technique for unlicensed channel access in WiFi is distributed coordination function (DCF). DCF uses binary exponential back-off with carrier sense multiple access/collision avoidance (CSMA/CA). When a UE has a data packet to transmit, it monitors the channel. UE transmits if the channel is found idle for a time equal to distributed interframe space (DIFS) time. Otherwise, if the channel is busy, UE waits for random backoff time, which is necessary to minimize chances of collision with data from other UEs. Moreover, UE has to wait for random backoff time between two consecutive data transmissions, even if the channel is found idle for DIFS time, which is necessary to avoid channel seizure by a single user. We assume that small cell while on the unlicensed band s_u , also maintains same random back-off mechanism, to ensure the coexistence with the WiFi A.P on unlicensed band w_u . According to author in [36], the probability to transmit for WiFi UE on unlicensed channel ρ_{w_u} is given by

$$\rho_{w_u} = \frac{2(1 - \rho_{wc})}{(1 - 2\rho_{wc})(\omega_w + 1) + \rho_{wc}\omega_w(1 - (2\rho_{wc})^\ell)}$$
(28)

where ρ_{wc} is collision probability, ℓ is a maximum back-off stage and ω_w is back-off window size. The collision probability for WiFi UE on unlicensed channel ρ_{wc} is given by

$$\rho_{wc} = 1 - (1 - \rho_{s_u})(1 - \rho_{w_u})^{n_w - 1}$$
⁽²⁹⁾

where ρ_{s_u} is the probability to transmit for small cell UE on the unlicensed channel, n_w is a number of users on WiFi. The probability of successful transmission by a WiFi UE is given by

$$S\rho_{w_{u}} = n_{w}\rho_{w_{u}} \left((1 - \rho_{w_{u}})^{n_{w}-1} \right) (1 - \rho_{s_{u}})$$
(30)

The probability to transmit for small cell UE on unlicensed band, $\,
ho_{\! s} \,$, is given by

$$\rho_{s_{u}} = \frac{2(1-\rho_{sc})}{(1-2\rho_{sc})(\omega_{s}+1) + \rho_{sc}\omega_{s}(1-(2\rho_{sc})^{\ell})}$$
(31)

where ρ_{sc} is collision probability, ℓ is a maximum back-off stage and ω_s is back-off window size. The collision probability for small cell UE on unlicensed channel, ρ_{sc} , is given by

$$\rho_{sc} = 1 - (1 - \rho_{w_u})^{n_w} \tag{32}$$

The probability of successful transmission by a small cell UE on unlicensed band is given by

$$S\rho_{s_{u}} = \rho_{s_{u}} \left(1 - \rho_{w_{u}}\right)^{n_{w}}$$
(33)

We assume that the unlicensed band is used for the same duration by the small cell and WiFi system and the small cell changes its minimum backoff window size adaptively. The time fraction occupied by the small cell on unlicensed band, can be defined as t_u , which will be equal to $S\rho_s$, i.e $t_u = S\rho_s$

. Based on the value of t_u , small cell adaptively changes the minimum size of back-off window. Small cell and WiFi will share unlicensed band, so we can write achievable sum rate for small cell on unlicensed band (R_{s_u}), as follows:



$$R_{s_{u}} = \sum_{k \in K} t_{u} \log \left(1 + \frac{p_{s_{u}}^{k} g_{s_{u}}^{k}}{N_{o} + \sum_{\substack{j=1\\j \neq k}}^{K} p_{w_{u}}^{j} f_{w,s_{u}}^{j,k} + p_{s_{u}}^{j} q_{s_{u}s_{u}}^{j,k}} \right)$$
(34)

`

where $f_{w,s_u}^{j,k}$ is channel gain between UE k on small cell s_u and UE j on WiFi A.P, $g_{s_l}^k$ is channel gain between UE and small cell licensed band and $q_{s_us_u}^{j,k}$ is channel gain between UE k on small cell s_u and UE j on another small cell s_u . Achievable sum rate for WiFi on unlicensed band (R_w).

$$R_{w_{u}} = \sum_{k \in K} \log \left(1 + \frac{p_{w_{u}}^{k} e_{w_{u}}^{k}}{N_{o} + \sum_{\substack{j=1\\j \neq k}}^{K} p_{s_{u}}^{j} f_{w,s_{u}}^{j,k} + p_{w_{u}}^{j} q_{w_{u}w_{u}}^{j,k}} \right)$$
(35)

where e_w^k is channel gain between UE and WiFi A.P on unlicensed band and $q_{w_u w_u}^{j,k}$ is channel gain between UE k on WiFi w_u and UE j on another WiFi w_u .

7.2.4 Achievable data rate in licensed spectrum

For simplicity we assume that macro cells and small cells use orthogonal channels in licensed band in TDM fashion with no interference, achievable sum rate for small cell on licensed band (R_{s_i}) is

$$R_{s_l} = \sum_{k \in K} t_l \log\left(1 + \frac{p_{s_l}^k g_{s_l}^k}{N_o}\right)$$
(36)

where t_l is time sharing factor for licensed band. Achievable sum rate for macro cell on licensed band R_m is given below

$$R_{m_l} = \sum_{k \in K} (1 - t_l) \log\left(1 + \frac{p_{m_l}^k h_{m_l}^k}{N_o}\right)$$
(37)

The optimization problem here is how to maximize sum rate of all UEs in network including *s* small cell UEs, *m* macro cell UEs and *w* WiFi UEs , by using power allocation and using parameters: $p_{s_u}^k, p_{w_u}^k, p_{m_l}^k, p_{s_l}^k$ and time allocation (using parameters: t_u, t_l), subject to minimum rate guarantee and co-channel interference threshold. Summary of notations is given in Table 23.

7.2.5 Problem formulation

,

We formulate joint power transmission and user association in licensed & unlicensed band such that sum rate is maximized. Mathematically, we have



- Constraint *C*1 is operator selection and BS is the selection constraint, which ensures that the UE is connected to one of the BSs (macro eNB on licensed band (m_l), small eNB on licensed band (s_l), small eNB on unlicensed band (s_u) and WiFi on unlicensed band(w_u) in one of the operators (operator A or operator B).
- *C*2 ensures that minimum rate requirement of each user is guaranteed.
- Constraint C3 to C6 is maximum power constraints for small eNB on the unlicensed band (s_u), WiFi on the unlicensed band(w_u), small eNB on licensed band (s_l) and macro eNB on the licensed band (m_l), respectively.
- C7 ensure that power experienced by any UE must be zero if it not connected to concerned BS. Constraints
- *C*8 and *C*9 guarantees the interference threshold. *C*10 is time to share constraint for the licensed and unlicensed band.
- *C*11 is minimum power constraint for each user.

7.2.6 Algorithm description

The combination of integer and continuous variables along with their non-linear behavior makes the problem in equation (34) very complex and challenging. However, by exploiting special structure of the problem, we can use mesh adaptive direct search algorithm (MADS) [37]to reach a sub-optimal solution. The MADS algorithm is an extension of generalized pattern search (GPS) algorithm [38].

The MADS is an iterative pattern search algorithm, it evaluates the objective function f on mesh of points. The mesh M_i at iteration i, given by

$$M_i = \bigcup_{y \in T_i} \{ y + \Delta_i^l \mathbf{D}_w : w \in N^n \},$$
(39)

where $\Delta_i^l \in \mathbb{R}^+$ is the size of mesh, T_i is set of points where the objective function is calculated at iteration i and $D \in \mathbb{R}^n$ is set of directions having a maximum of n_{Max} directions. D can be considered as $n \times n_{Max}$ a matrix containing n_{Max} directions. D must be positive spanning set [38], equal to the product of G and W (D = GW), where G is $n \times n_a$ nonsingular matrix and W is $n \times n_{Max}$ a matrix.

There are three main steps of MADS algorithm: search, poll, and update. In search step, the objective function f is evaluated at any finite set of points T_i on mesh in the feasible region. If the point y is not in the feasible region then the value of the function is set to $+\infty$. The search step allows the creation of point anywhere on the mesh, this flexibility restricts the search step to take part in convergence analysis. If an improved mesh point (y_{i+1}) is found, iteration may continue with search step or it may stop according to user's choice. If improved mesh point is not generated in search step, the poll step is invoked. The poll step explores space of optimization variable near current solution with following set of poll trial points

$$P_i = \{ y_i + \Delta_i^l d : d \in D_i \} \subset M_i$$

$$\tag{40}$$

SpeedSS

where D_i is positive spanning set, depicting poll directions. Points of P_i are generated so that their distance to the current solution y_i is limited by a parameter, called poll size $\Delta_i^p \in R^+$. Δ_i^p is always greater than $\Delta_i^l d$ (i.e $\Delta_i^p \ge \Delta_i^l d$). Moreover $\lim_{i \in I} \Delta_i^l d = 0$ if and only if $\lim_{i \in I} \Delta_i^p = 0$ for an infinite subset of iteration I. The update step determines whether iteration i was successful or not. This step updates parameters, Δ_i^l , Δ_i^p , T_i and at end of each iteration as shown in Figure 46.

7.2.7 Algorithm pseudo-code

The pseudo-code for MADS is given in the following Algorithm:

Mesh Adaptive Direct Search

// Initialization

1) $i \leftarrow 0$

2)
$$\Delta_o^l \in R^+$$
 , $\Delta_o^p \in R^+$, $y_o \in T_o$

// Search and Poll

3) The Search Step: Evaluate objective and constraint functions on finite number of points of $M(i, \Delta_i^l)$, to find y_{i+1}

4) The Poll Step: If y_{i+1} is not found, compute p MADS directions $D_i \in \mathbb{R}^n$. Construct a set of points $P_i \subset M(i, \Delta_i^l)$ with y_i , D_i , Δ_i^p . Evaluate the objective and constraint functions on p points of P_i

// Update

- 5) Determine success/failure of iteration i
- 6) Update solution(y_{i+1})
- 7) Update mesh(Δ_{i+1}^l)
- 8) Update poll size (Δ_{i+1}^p)
- 9) $i \leftarrow i+1$ Check stopping conditions and go to Search and Poll step.



7.2.8 Algorithm flowchart



Figure 46: Algorithm flowchart

7.2.9 Inputs and outputs

Inputs					
Parameter name/ID	Description	Provided by CRRM INTERNAL block	INTERNAL interface name/ID	Provided by an EXTERNAL block	EXTERNAL interface name/ID
k	Number of users	-	-	Yes	C_5G-X2AP
CQI	CQI-Update on Licensed and Unlicensed Bands			РНҮ	M_PHY_HMAC SAP
Sensing	Check if channel is busy using LBT			РНҮ	M_PHY_HMAC SAP
B_{w}	Total bandwidth on unlicensed band	-	-	Yes	C_SM-cRRM SAP
B_l	Total bandwidth on licensed band				C_SM-cRRM SAP
R^k	Min required rate				
С	Channels	-	-	Yes	C_SM-cRRM SAP

The Input and output variables that the proposed algorithm will use and produce are illustrated in Table 24 and Table 25 below:

Table 24:	Input parame	eters of prop	osed algorithm
-----------	--------------	---------------	----------------

Outputs					
Parameter name/ID	Description	Supplied to CRRM INTERNAL block	INTERNAL interface name/ID	Supplied to an EXTERNAL block	EXTERNAL interface name/ID
Selected Channels	The channel that each node is	-	-	Yes	M_cRRM_ConfigSAP



	assigned to in order				
	to transmit				
Data rate		-	-	Yes	M_KPI_cRRMSAP
Traffic load				Yes	M_KPI_cRRMSAP

Table 25: Output parameters of proposed algorithm

7.2.9.1 Mapping of proposed algorithms to the CRRM/MAC blocks

In Figure 47 the mapping of the algorithm is illustrated. Specifically, it can be seen the connection of the mechanism to the RRM functional blocks, RAT/Spectrum Selection Channel Selection and Traffic steering. Those blocks, of course, are connected with the Spectrum manager and Physical layer that provides the appropriate data of the Spectrum utilization as an input to the system and CQI information on each band (licensed/unlicensed). When the offloading decision is taken, this input is transferred by the cRRM to the higher MAC, where the MAC controller is in charge of selecting the appropriate unlicensed channel to sense according to the history of the unlicensed channels (i.e., spectrum availability and measured SINR). Then, the listen-before-talk (LBT) procedure is implemented and if the selected channel is sensed as available, the secondary carrier is configured and the related stream is scheduled by the MAC functions. On the contrary, when the channel is sensed as busy, the transmission is denied, and the MAC controller adjusts the cRRM offloading decision accordingly. Also when running in a more centralized way the spectrum manager will provide our algorithm with the operator spectrum ownership, the current spectrum allocations and even estimations of channel quality or occupancy at a geographical location. In addition, various data about the cells will be produced at the 5G-CELL block and supplied to the RRM blocks by the utilization of the C_5G-X2AP interface as shown. Likewise, the algorithm will be able to connect with the KPI collector where it can retrieve or send information about a specific cell or a range of cells of an area. Finally, the RRM algorithm will be able to exchange information through a two-way communication with the 5G MAC layer. More specifically two interfaces, the S_HMAC_cRRMSAP will provide inputs to the RRM blocks that will be used by the algorithm and the M cRRM ConfigSAP interface will support the information provided by the RRM to the MAC interface for configuration or reconfiguration when appropriate and required/requested by the algorithm.



Figure 47: Mapping to RRM functional blocks as defined in SPEED-5G D4.1.



Figure 48 illustrates a chart of the messaging sequence of the proposed mechanism. The algorithm will run in a centralized way. From there, information about the availability of RAT/Spectrum/Channels will be gathered. After a successful selection or even prediction of a particular channel has been completed, all the appropriate information and specifications will be requested and received to and from the MAC layer. At the MAC layer scheduling and inter-RAT coordination mechanisms will be enabled and run if necessary. Then, the physical layer will be reconfigured to the new RAT, band, and channel. Finally, cRRM will be achieved desired KPIs or in general the thresholds required by the system for optimal results. All channel selection are done by cRRM and at that time, the MAC is called to perform the scheduling and inter-RAT coordination.



Figure 48: Message sequence chart

7.2.10 Performance measures/KPIs

The performance measures and KPIs that will be used in the implementation of the proposed algorithm are introduced and explained in the below Table 26.

	Requirements			
КРІ	Massive IoT communications	Evolved Mobile Broadband		
User Experienced Data	From tens to hundreds of Khns	DL: 300 Mbps		
ate		UL: 50 Mbps		

Table 26: Performance measurements and KPIs of proposed algorithm

7.2.11 Simulation assumptions and parameters

Simulation parameters are presented in Table 27 and Table 28

Parameter	Value
Layout	21 cell Marco layout
ISD	500 m
System bandwidth	20 MHz (share between WiFi and small cells)
Carrier Frequency on unlicensed	5.8 GHz
Tx Power on unlicensed for LTE small cell and	24 and 30 dBm outdoor



Parameter	Value
WiFi AP	
UE noise figure	9 dB
Distance-dependent	Outdoor: (Smallcell-to- UE: ITU UMi)
pathloss/Shadowing/Fading	(WiFi-to- UE: ITU UMi)
	(UE-to- UE: 3GPP TR 36.843)
Number of Small cells	5 (2 indoor & 3 outdoor)
Number of users	10 per cell
Traffic model	3GPP Traffic-2
UE speed	3km/h
Scheduling	Proportional Fairness
SINR w.r.t CQI	(1.95, 4,6,8,11,14,17,19,21,23,25,27,29)
Frame	TDD
User association	The user will always be associated with a licensed layer (small cell), i.e., a user is associated to an SC over unlicensed band if it is also associated over licensed to the same small cell over licensed band. If the user is associated with small cell licensed layer, it can receive Wi-Fi or LTE-U (assume always in coverage).
LTE-U small cell dropping	Operators dropped randomly with min. the distance of 20m between small cells of the same operator.

Table 27: Simulation parameter LTE-U

Operator WiFi Parameter

Parameter	Value		
WiFi Type	802.11n (40 MHz)		
Number of WiFi AP	10		
Number of users	5		
MPDU	1500B and 1 ms duration		
	DCF		
MAC	Contention window: 15slot, Max:1023 slot		
	Detection: Energy detection		
CCA Threshold	-62dBm		
Channel rate	(13, 26, 39, 52, 78, 104, 117, 130) Mpbs		
Required SINR	(5,7,9,13,17,20,22,23) dB		
ACK frame rate	Max(6.5, 13, 26) Mbps <= Ch_Rate		
WiFi user Tx Power	18dBm		
ACK	16 Bytes		

Table 28: Operator WiFi Parameter



7.2.12 Performance evaluation

Figure 49 show the DL average user data rate vs. served traffic for different traffic load, when two LTE-U networks, operator A, and operator B coexist. Normally, the average user data rate drops when traffic increases, due to higher interference, longer contention, and scheduling delay. First, both operator A and B are using same energy detection thresholds (-62 dBm) and afterward we decrease the energy detection level to -72 dBm for both operators. Different energy detection has significant impact on the LTE-U performance. Having and ED threshold that is too low substantially reduces LTE-U channel access probability at medium and high loads. A higher LTE-U ED threshold leads to improved LTE-U performance, while WiFi performance remains higher than the reference case. This issue is caused by WiFi aggressive behavior against other technologies. To split fair channel access with WiFi, LTE-U should use an ED threshold similar to that employed by WiFi or adapt an ED threshold according to the scenario.



Figure 49: (a) Downlink user data rate of each network over total served traffic per AP, (b) 5th percentile SINR of operator A with two energy-detection thresholds



Figure 50: Ideal and non-ideal fronthaul

Figure 50 shows the average throughput of the system with and without ideal Fronthhaul (FH). For this simulation, we consider 100 cellular users (CUs). We also assume a fixed resource allocation between smallcell and marcocell. There are 100 resource blocks (RBs) in LTE 20 MHz band, which were divided equally among smallcell and marcocell users (50 RBs each). Each of these RBs is then assigned to their corresponding users via a proportional fairness scheduler.



When ideal fronthaul is deployed, the average throughput of the system is around 80, 70 and 64 Mb/s in low, medium and high traffic scenarios. Moreover, if we share resource allocation schemes between macrocell and smallcell with some interference cancellation mechanism are considered, the average throughput of the system increases even further.

When simulations with the non-ideal FH are run, a delay of 10 ms is experienced (see appendix A) and throughput drops to around 68,56 and 35 Mb/s in low, medium and high traffic scenario.



8 Conclusions

This deliverable finalized the design of RRM framework for SPEED-5G. The RRM framework comprises the centralized and distributed RRM and an adaptation layer that interfaces southbound to the upper MAC layer and northbound to OAM, spectrum manager, and KPI collector.

The benefits from this novel RRM framework are:

- Decoupling from algorithms via an abstraction layer, which allows algorithms to be added or changed in a modular fashion,
- The algorithms can operate centralised or distributed (or both),
- Support for multiple interfaces transparently from the algorithms point of view,
- Provision of a container for any data provided, or required by, the algorithms,
- Easily adapted to virtualization,
- Asynchronous procedures are enabled,
- The system is fully configurable by OAM/OSS.

Moreover, this deliverable also proposed the concept of *demulator* which is a piece of software combined with *de*monstrator and *emulator*. Its intended function is to (1) demonstrate the working of the RRM algorithms in several scenarios closely connected to the SPEED-5G use-cases, and (2) incorporate emulation code for RRM functions which can eventually be used in the real system. The demulator thus becomes in effect a *reference model* for the RRM and can be considered to represent a detailed design or definition of its algorithms.

Algorithms and initial simulation results are presented by different partners in the consortium. Detailed simulation results will be presented in D4.3.

Algorithm 1 is designed for efficient licensed-assisted access (LAA) operation in small cells, based on reinforcement learning. Running in the dRRM, the algorithm chooses the best-unlicensed channels to use on the downlink based on spectrum availability and the QoS requirements from the cRRM. Simulations show that there is an improvement in the worst served UEs.

Algorithm 2 is used for RAT/spectrum/channel selection based on hierarchical machine learning. Using dRRM it chooses the best option for the downlink taking into account a pool of bands and various licensing schemes (licensed/unlicensed/lightly-licensed) and the need to fulfil certain traffic requirements. The focus in the lightly licensed band of 3.5 GHz spectrum .

Algorithm 3 is Radio resource allocation with aggregation for mixed traffic in a WiFi coexisted heterogeneous network. It performs load balancing across WiFi and licensed spectrum. Using knowledge of the available capacity on the unlicensed spectrum it decides which UEs can use WiFi.

Algorithm 4 is a Fuzzy MADM strategy for spectrum management in multi-RAT environments. Working on the uplink, a connection manager (CM) is introduced on the UE side to collect the various components of the context and acts according to a policy that is remotely adjusted by the network manager. Based on this, a fuzzy multiple attribute decision making (MADM) implementation of the CM is developed to select the best RAT for a set of heterogeneous applications.

Algorithm 5 is Co-primary spectrum sharing in uplink SC-FDMA networks. The algorithm takes into consideration the users' buffer status and real-time delay constraints, as well as the operator priorities and the constraints of a realistic LTE system in order to perform uplink resource allocation in a QoS and energy efficient manner.

Algorithm 6 is Dynamic resource allocation algorithm for the coexistence of LTE-U and WiFi. The algorithm maximizes network throughput in the multi-operator scenario for 5G mobile systems by jointly considering a licensed & unlicensed band, user association and power allocation subject to minimum rate guarantee and co-channel interference threshold.



The next step is to select the most promising RRM technologies proposed in this document for further development and analysis, and the choosing of a small number of aspects to be carried forward to hardware in the loop demonstrations later in the project.



References

- [1] SPEED-5G, MS5 Enhanced system architecture, Oct 2015
- [2] SPEED-5G, D4.1 Metric definition and preliminary strategies and algorithms for RM, June 2016.
- [3] TS 36.331. Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification.
- [4] SPEED-5G, D5.1 MAC approaches with FBMC and simulation results, June 2016.
- [5] 3GPP Technical Specification Group Radio Access Network, "TR 36.889, Study on Licensed-Assisted Access to Unlicensed Spectrum, (Release 13)," V13.0.0, June 2015.
- [6] E. Kaufmann, O. Cappé, and A. Garivier, "On Bayesian Upper Confidence Bounds for Bandit Problems," Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS-12), pp. 592-600, 2012.
- [7] A. Anandkumar, N. Michael, A. K. Tang and A. Swami, "Distributed Algorithms for Learning and Cognitive Medium Access with Logarithmic Regret," in IEEE Journal on Selected Areas in Communications, vol. 29, no. 4, pp. 731-745, April 2011
- [8] V. Erceg, F. Baboescu, and T. Berham, "Discussion on LTE-WLAN Aggregation (LWA) and LTE WLAN Radio Level Integration with IPsec tunnel (LWIP)," IEEE 802.11-16/0437r1, March. 2016
- [9] S. Vassaki, A. Georgakopoulos, F. Miatton, K. Tsagkaris, and P. Demestichas, "Interference and QoS aware channel segregation for heterogeneous networks: A preliminary study," 2016 European Conference on Networks and Communications (EuCNC), Athens, June 2016, pp. 195-199
- [10]Ericsson, "Ericsson mobility report on the pulse of the networked society," White Paper, Nov. 2015. [Online] Available: https://www.ericsson.com/res/docs/2015/mobility-report/ ericssonmobility-report-nov-2015.pdf
- [11]Nokia Corporation, NTT Docomo Inc., and Nokia Networks, "LTE Carrier Aggregation Enhancement Beyond 5 Carriers," 3GPP Work Item Description (RP-142286), Dec. 2014
- [12]V. Erceg, F. Baboescu, and T. Berham, "Discussion on LTE-WLAN Aggregation (LWA) and LTE WLAN Radio Level Integration with IPsec tunnel (LWIP)," IEEE 802.11-16/0437r1, March. 2016
- [13]4G America, LTE Aggregation & Unlicensed Spectrum, Nov. 2015
- [14] Netmanis, Analysis of LTE-WiFi Aggregation Solutions, technical report, March, 2016
- [15]H. Lee, S. Vahid, and K. Moessner, "A Survey of Radio Resource Management for Spectrum Aggregation in LTE-Advanced,", IEEE Communications Surveys and Tutorials, vol. 16, no.2, pp. 745-650, Second Quarter 2014.
- [16]A. Abdelhadi and C. Clancy, "An optimal resource allocation with joint carrier aggregation in 4G-LTE," 2015 Int'l Conf. on Computing, Networking and Communications (ICNC), pp. 138-142, 2015.
- [17]S. Boyd, Convex Optimization, Cambridge University Press, 2004
- [18]N. Chen, P. Hu, X. Tao, Q. Cui and Y. Shang, "Achievable energy efficiency resource allocation in OFDM system with mixed traffic," 2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC), Washington DC, 2014, pp. 1046-1050.
- [19]David Martin, "3GPP Long Term Evolution: Performance Analysis and Evolution towards 4G with Coordinated Multi-Point Transmission", Feb 2016.
- [20]3GPP TR 36.814, Further Advancements for E-UTRA physical layer aspects , March 2010
- [21]J. A. Flores, D. Robalo and F. J. Velez, "Transmitted power formulation for the implementation of spectrum aggregation in LTE-A over 800 MHz and 2 GHz frequency bands," 2013 16th Int'l Symp.



Wireless Personal Multimedia Communications (WPMC), Atlantic City, NJ, 2013, pp. 1-5

- [22]F. Zhou, L. Feng, P. Yu and W. Li, "Energy-efficiency driven load balancing strategy in LTE-WiFi interworking heterogeneous networks," 2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), New Orleans, LA, 2015, pp. 276-281.
- [23] "5G Whitepaper: The Flat Distributed Cloud (FDC) 5G Architecture Revolution," Institute for Communication Systems, 5G Innovation Centre, University of Surrey, Tech. Rep., January 2016.
- [24] "Simulation assumptions and parameters for FDD HeNB RF requirements," 3GPP, Tech. Rep. 3GPP R4-092042, May 2009.
- [25] "NS-3 Model Library," Tech. Rep. Release NS-3.26, March 2016. [Online]. Available: https://www.nsnam.org/docs/release/3.26/models/ns-3-model-library.pdf.
- [26]F. Bouali, K. Moessner, and M. Fitch, "A Context-aware User-driven Framework for Network Selection in 5G multi-RAT Environments," in IEEE 84th Vehicular Technology Conference (VTC Fall), September 2016
- [27] "OMA Device Management Protocol," Open Mobile Alliance, Tech. Rep. version 1.2.1, June 2008.
- [28]T. J. Ross, Fuzzy logic with engineering applications. Chichester, U.K. John Wiley, 2010.
- [29]"Study on licensed-assisted access to unlicensed spectrum, (Release 12)," 3GPP, Tech. Rep. 36.889-V13.0.0, June 2015.
- [30]3GPP TS 36.213, V12.4.0 (2017-03): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures, Rel. 14.
- [31]R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," DEC Research Report TR-301, Sept. 1984.
- [32]D. Triantafyllopoulou, T. Guo, and K. Moessner, "Energy Efficient ANDSF-assisted Network Discovery for non-3GPP Access Networks," in *Proc. IEEE Int. Workshop CAMAD Commun. Links Netw.*, Barcelona, Spain, 2012, pp. 297-301.
- [33]H. Holma, and A. Toskala, LTE for UMTS: OFDMA and SC-FDMA based radio access. Chichester, U.K.: John Wiley & Sons, 2009.
- [34] http://www.hhi.fraunhofer.de/de/kompetenzfelder/imageprocessing/research-groups/imagevideo-coding/svc-extension-ofh264avc/jsvm-reference-software.html.
- [35]Qualcomm white paper, "LTE in Unlicensed Spectrum: Harmonious Coexistence with WiFi," https://www.qualcomm.com/media/documents/files/extending-the-benefits-of-lte-advanced-to-unlicensed-spectrum.pdf, Jun. 2014.
- [36]G. Bianchi. Performance analysis of the ieee 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000.
- [37]Charles Audet and John E Dennis Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on optimization*, 17(1):188–217, 2006.
- [38]Virginia Torczon. On the convergence of pattern search algorithms. *SIAM Journal on* optimization, 7(1):1–25, 1997.



Appendix A Fronthaul/backhaul requirements for small cells when deployed on the SPEED-5G use-cases

The meaning of fronthaul and backhaul is illustrated in Figure 51 below. Fronthaul is the connection of the remote radio head (RRH) to the Baseband unit (BBU), and backhaul is the connection of the BBU to the core network. To avoid over-complication we do not venture to describe 'midhaul' or any other variant.



Figure 51: Definition of fronthaul and backhaul

For RRHs within homes, the fronthaul will typically consist of the broadband connection to the home, which often transitions a street cabinet, and a further link to a telephone exchange or similar building owned by the network operator. The BBU(s) can be housed in this building. With Fibre to the Cabinet (FTTC), the first leg of the broadband link, from the house to the cabinet, is likely to be xDSL or G.FAST copper technology, with fibre between the cabinet and the exchange building.

We now come to the question of typical delays in the fronthaul and backhaul on a typical network. If the fronthaul has a copper segment using ADSL2, the latency is around 25ms each way for the DSL modem, plus 0.5ms for each 100km that the data travels. If the copper segment is G.FAST, the minimum delay is determined by the frame rate which is 750µs, but this is increased by any coding and HARQ mechanisms and typically the delay is a minimum of 2ms each way. It is possible to dispense with the HARQ mechanism on G.FAST to reduce the fronthaul latency but this is not always appropriate if the protocol stack is sensitive to bit-errors on the fronthaul.

The backhaul section in Figure 51 is typically 1ms per MPLS hop, and there could be several hops in the backhaul connecting the BBUs to the mobile core. We should budget for 10ms in the backhaul as shown in Table 29.

Section	Fronthaul	Backhaul
Latency range	>25ms each way if ADSL2 is	Up to 10ms
	used,	
	750µs to >2ms each way if	
	G.FAST is used, depending on	
	HARQ	

Table 29: Latency assumptions on fronthaul and backhaul

A.1 Location of the fronthaul

We now consider where to place the fronthaul in the small cell stack. Figure 52 is based on the SPEED-5G small cell architecture design, and shows where the stack mechanisms are located.



A.1.1 Below the PHY layer



Figure 52: SPEED-5G small cell stack with fronthaul below PHY layer

Note that the scheduler works over the lower three layers (RLC, MAC and PHY). The MAC layer contains the HARQ mechanism, the RLC performs ARQ and is responsible for maintenance of logical channels across the air interface. Above the RLC layer is the PDCP, whose task is to perform header compression, encryption and also to check and correct the packet ordering.

Figure 52 shows the fronthaul at the bottom of the stack, so that the IQ samples from the PHY layer processes are transferred to / from the RRH for modulation / demodulation and up-conversion/ down-conversion. Putting the fronthaul in this position has the following implications for the fronthaul:

Cons:

- A huge overhead. A very high bit-rate is needed, which is multiplied for multiple antennas and can easily reach several Gbit/s even for a moderately low bandwidth channel,
- It needs to be very low latency to maintain proper multiple antenna operation and not to upset frame timing ,
- Need for time phase synch,
- LIPA needs further consideration and may not be usable.

Pros:

- Errors on the fronthaul can be assumed corrected by the *ARQ and FEC within the stacks or at the UE.
- The RRH is very simple and low-cost, and most of the small cell processing and complexity is removed to the centralised BBU, consequently saving costs.

These requirements lead to the conclusion that a CIPRI-like fibre interface is needed if the fronthaul is in this position, which is not a good fit to typical deployments.

A.1.2 At the MAC layer

Here we show the fronthaul between the MAC and PHY layers as shown in Figure 53.

Putting the fronthaul in this position means that some scheduling is performed in the RRH and some in the BBU. Frames for transmission are transferred across the fronthaul, with little additional overhead. The error-correcting mechanisms in the RLC and MAC layers can be used to combat errors in the fronthaul, although the effectiveness of these mechanisms against the particular types of error



on copper fronthaul will need further investigation. Putting the fronthaul in this position has the following implications for the fronthaul:



Figure 53: Fronthaul between MAC and PHY

Cons:

- The round-trip delay must be within 8ms (in the case of LTE) for UE network entry
- Uplink HARQ cannot be used, so ACKS must be spoofed and then any errors on the uplink need to be mitigated through re-transmissions, which reduces the efficiency
- LIPA needs further consideration and may be unusable.

Pros:

- Little additional overhead to the cell load (<10%),
- Centralised scheduling means more concurrent users, which may be beneficial for MTC
- Significant processing has been removed from the home equipment with consequential simplification and cost saving

A.1.3 At the PDCP layer

The final position we consider for the fronthaul is below the PDCP layer as shown in Figure 54.



Figure 54: Fronthaul between PDCP and RLC



Putting the fronthaul in this position means that all scheduling is performed in the RRH and name in the BBU. Packets for transmission and reception are transferred across the fronthaul, with little additional overhead. The error-correcting mechanisms in the RLC and MAC layers are now not available to combat errors in the fronthaul, and error propagation is a topic for further work.

Putting the fronthaul in this position has the following implications for the fronthaul:

Cons:

- The simplification of the home equipment, and hence cost savings, is not as significant as placing the fronthaul in other locations
- No correction of errors that occur in the fronthaul,

Pros:

- Not sensitive to latency (so that error correction can be carried out in the fronthaul if it is G.FAST)
- Little overhead is added to the traffic being carried by the cell (<10%)

A summary of fronthaul location options is shown in Table 30.

Split point	Pros	Cons
РНҮ	Potential to integrate backhaul into channel	Large overhead, need for high bit- rate, low latency and time phase synch. Needs point to point fibre links
MAC	Synergy through centralising some scheduling. Small overhead, low sensitivity to latency	Need to spoof UL acks if latency is >8ms which may limit UL speeds. No correction of errors on UL, so re- transmissions need re-scheduling
PDCP	Small overhead, low sensitivity to latency	No correction of errors that happen in fronthaul on UL or DL, and re- transmissions will need re- scheduling

Increasing cost benefit to RRH

Table 30: Summary of fronthaul location pros and cons

A.2 Practical measurements

Since we can discount having the fronthaul below the PHY layer because of the need for very high CIPRI-like bit-rates and also of the need for tight time phase synchronisation, we decided to focus the lab testing on the other two options which is the where the fronthaul is located between the RLC / MAC layer and also between the PDCP / RLC layer. The BBUs and RRHs equipment used for the testing was supplied by Cavium, and the G.FAST equipment was supplied by Huawei. The radio layer used for these tests was LTE.

A.2.1 Test 1 – Direct Ethernet connection

We installed Ethernet interfaces to the layers, in collaboration with the suppliers, for both RLC and MAC/PHY splits. Figure 55 shows the test equipment layout and the latency results.



	BBU-UE Latency mi	n/avg/max: 13/25/40ms		
RLC Split ≈ 90Mbps ←	BBU-RRH Latency min/a		Media Server	
REL		AND		
	Ethernet	Ethernet	Ethernet	BBU/EPC/APP
MAC/PHY Split		Switch		HUNDERX
≈ 60Mbps				
<				>

Figure 55: Test equipment layout and latency results for direct Ethernet connection

The bit-rate overheads associated with the above fronthaul splits are 10% and 16% for RLC and MAC/PHY splits respectively.

A.2.2 Test 2 – With G.FAST

We extended the test to include a 100m length of typical broadband copper pair cable running G.FAST, using a pair of Ethernet to G.FAST modems, as shown in Figure 56. This is representative of many broadband connections to homes.



BBU-UE Latency min/avg/max: 19/30/46ms

Figure 56: Extension of test using 100m of copper pair running G.FAST

Note that these latencies are all one-way. From these tests it can be seen that the fronthaul latency over 100m of G.FAST including Ethernet modems is 3ms maximum. It can also be seen that the LTE air interface adds a significant amount of latency, of the order of 10 - 20ms on average. It is also evident that there is a significant variation in the latency in all cases, with a maximum to minimum ratio of greater than 2:1. So, we can conclude that 5G radio needs to perform 10 to 20 times better than LTE to achieve 1ms radio latency.

The fronthaul latency is below 3ms, which is consistent with the G.FAST assumptions in table 1. In the actual test, no HARQ is used on the G.FAST so one might expect the delay to be 750µs, but there are 1 to 2ms delays in the Ethernet to G.FAST converters which we assume would not be present in a practical deployment.

Splitting at the RLC layer is less susceptible to timing delay and jitter than the MAC/PHY layer. When we tried 300m of G.FAST, the MAC/PHY split started to have throughput problems because it typically requires a one-way delay of 3 - 10ms. We therefore have a trade off as expected; more cost savings are obtained by putting the fronthaul at the MAC/PHY layer, but this is more sensitive to delay and delay jitter and it also has a higher overhead (16% as opposed to 10% at the RLC layer).