Management and Orchestration Architecture for Integrated Access of Satellite and Terrestrial in 5G

Taesang Choi
Intelligent Network Research Lab
ETRI
Daejoen, Rep. of Korea
choits@etri.re.kr

Seok Ho Won
Future Mobile Communication Lab
ETRI
Daejoen, Rep. of Korea
shwon@etri.re.kr

Alessandro Giuseppi,
Dep. of Computer, Control and Management Engineering "Antonio Ruberti"
University of Rome "La Sapienza" and Space Research Group CRAT
giuseppi@diag.uniroma1.it

Antonio Pietrabissa
Dep. of Computer, Control and Management Engineering "Antonio Ruberti"
University of Rome "La Sapienza" and Space Research Group CRAT
pietrabissa@diag.uniroma1.it

Sungoh Kwon
School of Electrical Engineering
University of Ulsan, Ulsan, Korea
sungoh@ulsan.ac.kr

***Abstract*— Multi-RAT access network, or heterogeneous access network, is considered to be the key enabling technology to satisfy the 5G requirements, such as high data rate, ultra-low latency and reliability. To make efficient use of all the available network resources, various research activities on multi-connectivity have been proposed to simultaneously connect, steer, and orchestrate across multiple different radio access technologies. Standardization of the management and orchestration of multi-connectivity environment, however, has just been initiated, thus further research and development is required. This paper proposes a novel management and orchestration architecture for integrated access of satellite and terrestrial in 5G. It especially focuses on the traffic steering and load-balancing of heterogeneous multi-RAT access environment.**

Keywords—multi-connectivity, load-balancing, management and orchestration, traffic steering, and QoS/QoE management

I. Introduction

Multi-RAT access network, or heterogeneous access network, is considered to be the key enabling technology to satisfy the 5G requirements, such as high data rate, ultra-low latency and reliability. To make efficient use of all the available network resources, multi-connectivity has been proposed to simultaneously connect, steer, and orchestrate across multiple different radio access technologies. The main advantage of the multi-connectivity approach fosters the possibility to send the user traffic in different Radio Access Technologies (RATs) that better satisfy the service requirements and the user needs. Some of the important advantages that can achieved by multi-connectivity are:

- Improvement of the overall data rates (throughput) to mobile UEs in 5G networks for both DL and UL.
- Improvement of the optimal exploitation of 5G network resources while meeting the 5G KPIs.
- Guarantee of service continuity (reliability) in mobile UEs

They, however, do not just come by without efforts of appropriate management and orchestration of multi-RAT environment. The main challenge is full lifecycle management of resources involved in the multi-connectivity environment: resource status observation, analytics, decision, and execution. Observation of precise resource status of both satellite and terrestrial RAT and that of the associated core network requires new capabilities across UE, gNB, and CN. Standardization of interfaces, protocols and mechanism of them are also essential for global interoperability. Analytics plays a very important role to accurately diagnosis optimal resource usage in multi-connectivity environment. Based on the analytics results, intelligent decision making for the optimization of resource usage can be performed by utilizing AI-assisted load-balancing algorithms. Lastly, execution process can trigger traffic steering and load-balancing control and management actions which have been made during the decision making process.

ETSI 3GPP Release 15 5G architecture [1] has been standardized in September 2018 and more advanced capabilities including network slicing, network and management data analytics function, traffic steering, and QoS control and monitoring are under development as Release 16 targeted for its completion in March 2020. However, management and orchestration of multi-connectivity environment has just been initiated and will be completed in Release 17. Our research, 5G All-Star (5G AgiLe and fLexible integration of SaTellite And cellulaR) [2], is trying to define multi-connectivity management and orchestration architecture and develop associated enabling technologies by analyzing the gaps present in the current standards and available technical solutions.

In this paper, we propose a novel architecture for management and orchestration specifically for the multi-connectivity environment in 5G. The paper is organized as following. Section II describes core technologies to enable optimal multi-connectivity architecture in 5G. Section III explains our proposed management and orchestration architecture for optimal resource usage in multi-RAT environment. Section IV provides a prototype implementation efforts with preliminary performance evaluation results. Section V summaries our ongoing research efforts with potential future work.

II. Enabling Technology

This section provides the state of the art concerning 5G networks related to the 5G multi-connectivity, its management and orchestration, and essential enabling technologies and standards.

978-1-7281-4199-2/20/$31.00 ©2020 IEEE

First enabling technology is multi-RAT integration and mangement methods. In [3] three multi-RAT integration methods are presented:

Application Layer Integration: it consists of a higher-layer interface, providing information exchange between UEs and content provider, over multiple RATs. This solution can be easily implemented, but it is an application-dependent and may not fully take into account the network state, which leads to suboptimal exploitation of resources, especially if the network state is observed to vary dynamically.

Core-Network-Based Integration: this solution is proposed by 3GPP for cellular/WLAN integration based on interworking between core networks. In this case, the RAT selection is made considering operators' policy for network selection, but the overall network selection decision remains in control of the UE. The UE is then able to take its decisions considering operator policies, radio links performances and user preferences. It is worth remarking that typically the UE only has local knowledge about the network conditions, resulting in the suboptimal selection of decisions, degrading the overall network performances and the Quality of Experience (QoE) of its user.

RAN-Based Integration: this solution is proposed by 3GPP in NR/LTE dual-connectivity and allows coordination between the RATs using dedicated interfaces. The cooperation level between the different RATs is constrained by the back-haul links. Having high backhaul link capacities allows full cooperation between RATs, enabling more dynamic RRM mechanism and improving overall system and user performances. In addition, the central units may be employed as a mobility and control anchor. The benefits of this solution are the adaptation of the decisions to dynamic variations in the radio links conditions, consequently minimizing session interruptions or packet drops. Furthermore, in this configuration, appropriate feedback from UEs and operator preferences can be considered in the RATs selection.

Furthermore, in [3], multi-connectivity management approaches are presented:

User-Centric Approach: with this solution the UE is continuously monitoring the radio links conditions, and, considering thresholds-based performance parameters (e.g. SNR), the RAT selection can be performed. In advanced scenarios, the UE can consider other RATs characteristics (e.g. coverage) to better satisfy the application and user needs.

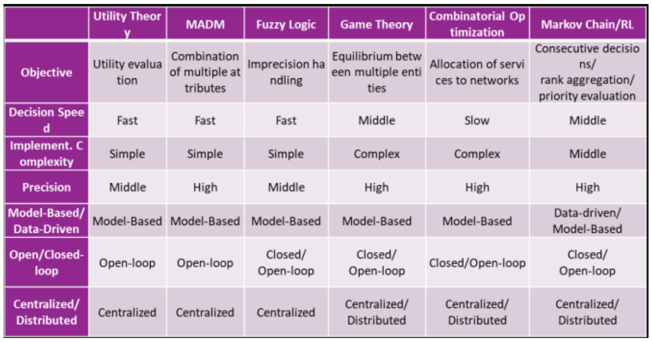
RAN-Assisted Approach: User-Centric approach is limited to the local UE knowledge. For instance, the UE performs RAT selection based typically on the SNR, and in a highly dense environment the selection decision typically doesn't remain effective for long, due to the varying load of the RATs. The RAN-assisted approach employs network assistance from the RAN to the UE for RAT selection decisions. An example of assistance parameters can be network load, RAT utilization, expected resources allocation, etc.

RAN-Controlled Approach: the above-mentioned schemes are user-centric by nature, resulting in suboptimal decisions from the overall system performances point of view. The RAN-controlled approach places the multi-connectivity control in the radio networks. In this approach the RAN can assign the UEs to certain RATs. Such a solution can be distributed across RATs or may utilize a central unit that manages radio resources across several cells/RATs. The UEs, in this solution, is configured to report radio measurements on their local radio environment. This solution is adopted by 3GPP for addressing dual-connectivity issues.

3GPP SA5 is recently defining a standard for "management and orchestration aspects with integrated satellite components in 5G network"[4]. It identifies the main key issues associated with business roles, service and network management and orchestration of a 5G network with integrated satellite component(s) (whether as NG-RAN or non-3GPP access, or for transport). As solutions, it defines a management architecture of integrated access of satellite and terrestrial RATs.

The second enabling technology is traffic steering and load-balancing algorithms to assist making optimal decision of RAT selection. The problem consists in the selection of the most appropriate access network with characteristics able to satisfy the 5G KPI requirements. These selections can be performed by considering different network features as for instance: the mobility of the network nodes, the QoS attributes, the energy constraints, etc.. The algorithms capable to perform the RAT selection are evaluated by considering the algorithms characteristics like computational complexity, implementation complexity, distributed or centralized deployment with either open or closed-loop type, dynamic or static behavior, model-based or data-driven. RAT selection approaches, already investigated in the literature, concern the use of mathematical theories with the main characteristics detailed in Table 1.

Table 1: Key Characteristics of Mathematical Theories for Network Selection



	Utility Theory	MADM	Fuzzy Logic	Game Theory	Combinatorial Optimization	Markov Chain/RL
Objective	Utility evaluation	Combination of multiple attributes	Imprecision handling	Equilibrium between multiple entities	Allocation of services to networks	Consecutive decisions/ rank aggregation/ priority evaluation
Decision Speed	Fast	Fast	Fast	Middle	Slow	Middle
Implement. Complexity	Simple	Simple	Simple	Complex	Complex	Middle
Precision	Middle	High	Middle	High	High	High
Model-Based/ Data-Driven	Model-Based	Model-Based	Model-Based	Model-Based	Model-Based	Data-driven/ Model-Based
Open/Closed-loop	Open-loop	Open-loop	Closed/ Open-loop	Closed/ Open-loop	Closed/Open-loop	Closed/ Open-loop
Centralized/ Distributed	Centralized	Centralized	Centralized	Centralized/ Distributed	Centralized/ Distributed	Centralized/ Distributed

Besides load-balancing algorithms, it is also important to provide network data and management data analytics function for generating accurate data that these algorithms can utilize for their decision making. 3GPP is defining related standards: NWDAF (Network Data Analytics Function) [5] and MDAS (Management Data Analytics Service) [6]

The third key enabling technology for optimal multi-connectivity management and orchestration is network slicing. Since it is important to share the limited physical satellite and radio resources, network slicing technology allows resource sharing by creating network slice per network operator and its specific KPI. 3GPP is currently defining various standards for network slice lifecycle management and orchestration including planning, commissioning, operation, and de-commissioning processes [7].

III. MANO ARCHIITECTURE FOR 5G MULTICONNECTIVTY

Based on the key enabling technology gap analysis, we defined a 5G-ALLSTAR multi-connectivity management and orchestration architecture as depicted in Figure 1.

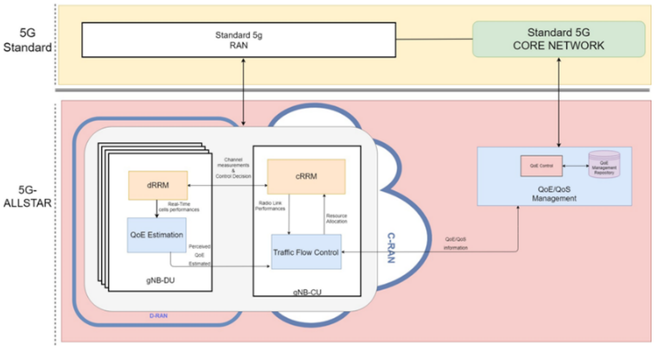


Figure 1. Multi-connectivity management and orchestration architecture

For the optimal distributed resource control and maximization of the whole user's experience, 5G-ALLSTAR project proposes to enrich the current 5G entities, i.e., Core Network, Cloud Radio Access Network and Distributed Radio Access Networks, with advanced functionalities to satisfy and control specific end-to-end Services/Applications by using both Traffic Flow Control and Quality of Experience (QoE) Control.

The QoE/QoS Management is a module aimed at enhancing the standard QoS Control that 5G implements through the QoS Profile. Such an enhancement is performed by means of the so-called Connection Preferences which are deduced by the QoE/QoS Management taking into account personalized connection requirements aimed at satisfying even the subjective Quality of Experience of the user handling the connection in question; so, the Connection Preferences include QoE-related requirements which are additional with respect to the QoS-related requirements included in the 5G "standard" QoS Profile. The QoE/QoS Management functionality is logically distributed into the Core Networks (CNs) and 5G management system and includes:

The QoE Management Repository which includes information relevant to past and current connections managed by the CN. The repository stores the following information:

- Connection Id which identifies the key parameters of the connection; in particular, it includes the following subfields: (i) Source UE Id, (ii) Destination UE Id, (iii) Service Type, (iv) QoS Profile, (v) User Equipment (UE) type.
- Connection Preferences deduced, at each connection set-up, by the QoE Control by means of AI-based algorithms. The Connection Preferences are not modified for the whole connection duration;
- Connection QoE History, which includes, for each of the cells which has served the Connection (if already terminated), or is serving the Connection (if it is still in progress) the following subfields: (i) Cell-Id, (ii) Time Duration, (iii) Cell QoS Performance, (iv) Implicit QoE Feedbacks (i.e. feedbacks related to the Perceived QoE computed by a suitable QoE Estimation module which is provided to the QoS/QoE Management module by the Traffic Flow Control module. The Explicit QoE Feedbacks (i.e. feedbacks related to the Perceived QoE directly provided by the users involved in the Connection) is provided by the Content Providers directly to the QoE/QoS Management which stores such information in the repository. The updates of the Connection QoE History relevant to a given cell serving a given connection are provided by the Traffic Flow Control module to the QoS/QoE Management whenever the Cell in question no longer serves the Connection in question.

The QoE Control module, at each connection set-up, is in charge of deducing the Connection Preferences by analyzing (by means of suitable AI-based techniques) the (big) data included in the QoE Management Repository. The rationale of the Connection Preferences is to include personalized QoE-related requirements which are additional with respect to the QoS-related requirements which are associated with the standard 5G QoS Profile. At each connection set-up, the QoE Control (i) deduces the Connection Preferences, (ii) stores them in the QoE Management Repository, (iii) sends the Connection Preferences, related to a specific Connection and UE, to the gNB-CU and, in particular, both to the Traffic Flow Control module and to the cRRM module. At each Connection termination, the QoE Control has to inform the gNB-CU about the termination of the Connection.

Figure 2 shows the multi-connectivity management and orchestration functionality distribution in 5G network. Resource control including QoS/QoE control functionalities are distributed at gNB-DU, gNB-CU, and CN and its management and orchestration functionality is located at 5G management system. Real-time control is executed at network level, that is, in cooperation between gNBs and CN. Management decisions and actions are performed by the 5G management system. It includes both real-time, near real-time, and non-real-time management actions.

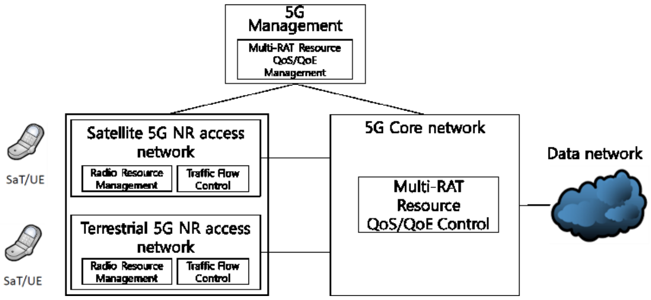


Figure 2. Multi-connectivity control and management relationships

For the optimal multi-connectivity management, the architectural selection of the functional splits, i.e., the decision about which function should be placed in either the central or the distributed units of gNB, is a crucial point that defines the whole traffic flow control system.

From the control plane perspective, the ideal scenario would be to put the whole set of functionalities that are technology independent as well as non-real-time and low bit rate, (e.g., traffic steering, spectrum sharing, etc…) in the central unit in order to have a complete view of the system, allowing optimal decision making. In this case, the distributed units have technology dependent, real-time and high bit rate functionalities in order to meet their requirements.

Regarding the protocol stack split there are three main options as shown in Figure 3:

- Intra-PHY split: in this case, there are the requirements of low latency (about 1 ms one-way delay) and high throughput in the fronthaul, but there is no requirement of high computing power in the distributed units;
- PHY-MAC split: in this case the throughput is reduced compared with the intra-PHY split, but the same latency constraints are present. In this case, the amount of computing power in the distributed units is higher than the previous case;
- PDCP split: this case is the most attractive for the relaxed latency requirements (tens of ms) with a throughput like the PHY-MAC, but there is a significant need for computing power in the distributed units.

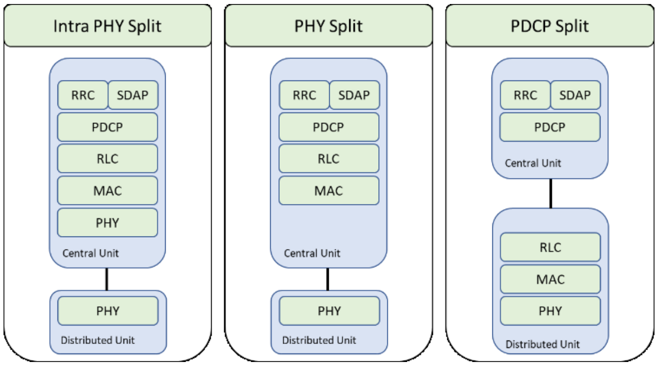


Figure 3. Functional Splits

5G-ALLSTAR choice is a common PDCP for the user plane and a common RRC for the control plane. In contrast to PHY, MAC and RLC functions, the PDCP functions do not have rigorous constraints in terms of synchronicity with the lower layers. Furthermore, this option will allow traffic aggregation, as it can facilitate the management of traffic load and this split has already been standardized for LTE Dual Connectivity [8].

For the verification of our PoC implementation for multi-connectivity, an example of Multi-Connectivity physical architecture with three RATs and two UEs in a downlink scenario is presented in 4.

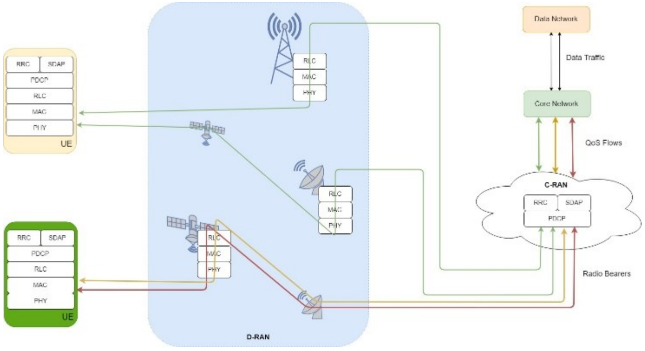


Figure 4. Multi-Connectivity Physical Architecture

The whole set of available RATs have common RRC and partially common UP (SDAP and PDCP layers). The RRC and PDCP common layers approach in the C-RAN bring several advantages i.e., the fast switch/UP aggregation and PDCP split.

The architecture in Figure 4 is composed of three RATs: i) a terrestrial BS; ii) a transparent satellite, and iii) a regenerative satellite. The figure presents two data flows (Data Traffic) coming from the Data Network. The Core Network divides the data flows in three QoS Flows by using the UPF functionalities (considering the SMF configuration) and, in turn, the C-RAN sends the QoS Flows to the UEs with a proper selection of radio bearers.

The C-RAN entails RRC functions capable to (i) configure the SDAP for the mapping of QoS Flows into data bearers; (ii) configure the other UP layers to establish the data bearers with the desired performances. These functionalities are performed by the cRRM and Traffic Flow Control module as defined in the description of Figure 1.

IV. PROTOTYPE IMPLEMENTATION

In this section, we describe our prototype implementation efforts based on the proposed architecture. We are currently developing various modules of multi-connectivity management and orchestration system including QoE/QoS personalization module, QoE/QoS traffic flow control module, QoE/QoS resource load-balancing algorithms, and overall management and orchestration module.

The Quality of Experience/Quality of Service Management is a fundamental part of a "Personalisation system", see Figure 5. The 5G-ALLSTAR Personalization System is aimed at providing a non-standardized Connection Preferences which enriches the inputs deployed to the Traffic Flow Control. These Connection Preferences will be considered during the Multi-Connectivity assignment tasks. The Connection Preferences contain a set of parameters deduced by the QoE Control module. The Personalization System is also able to estimate the perceived QoE for each on-going service/application at each UE by using the QoE Estimation module.

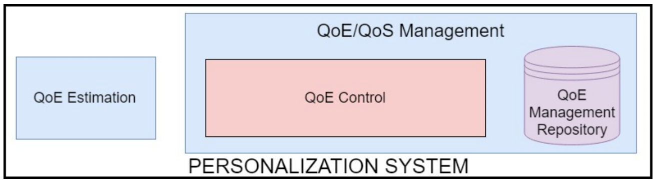


Figure 5. Personalization System

The 5G networks aim at satisfying at the same time a large diversity of UE requirements. This requires that the network be flexible and adaptable to different traffic types, as defined by the 5G requirements which need a granular approach to the QoS handling. It is known that each UE could have one or more PDU sessions which may have one or more QFI (QoS Flows Identifier) at the same time. In this respect, a granular assignment of the QoS markers into the PDU session/s for different traffic types is a fundamental feature in the design of a 5G network for both Downlink (DL) and Uplink (UL) cases. It also allows the Access Network to handle the data packets with different QFI, by assigning the different QoS flow to the most suitable data radio bearers.

The QoE Control module (see Figure 6), in the 5G-ALLSTAR project, is developed to assign for each PDU session the enriched QoS Profile needed to cope with the related QoS Requirements. The QoE Control module relies on the information stored into the QoE Management Repository and the service information provided by the data provider. The QoE Control module includes algorithms that will be able to produce the Connection Preferences identified for each UE and Connection by correlating different input data.

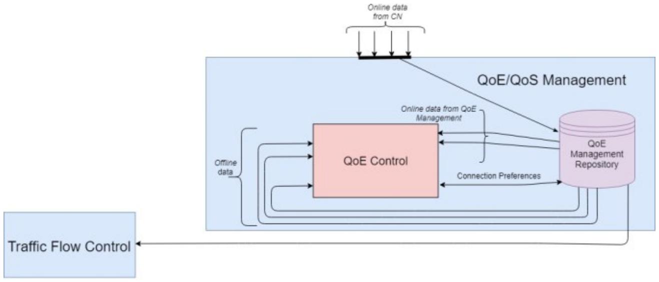


Figure 6. QoE Control Module Implementation Architecture

The 5G-ALLSTAR project is implementing Multi-Connectivity solutions to satisfy both (i) data rate boosting and (ii) service continuity for reliability purposes with the implementation of innovative control methodologies leveraging the already available tendency in the cloud-based solutions for improving network performances. As already described above, in 5G-ALLSTAR project the Traffic Flow Control module will be in charge of ensuring the 5G KPIs in terms of latency, data rate, reliability, etc, by combining different methodologies for delivering the suitable traffic steering, splitting and switching solutions in order to handle Multi-Connectivity mechanisms. The methodologies behind the Traffic Flow Control module will set up both SDAP layer and PDCP layer in the gNB-CU for enabling the Multi-Connectivity where the radio access points do not need to be divided in master node or secondary node but assuming to have the Control Plane functionalities in the C-RAN and the split functions are at the PDCP level.

The traffic flow control algorithm will be able to decide the dynamic association of the traffic of a given UE with one or more RATs. Indeed, these decisions will be based on the information about the traffic, the UE and the RATs conditions as shown in Figure 7.

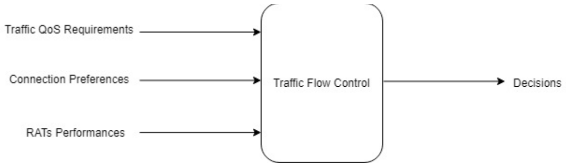


Figure 7. Traffic Flow Control Input / Output

As traffic flow control algorithm, we are developing a few load-balancing algorithms based on the gap analysis in Table 1. They are wardrop equilibrium 9], reinforcement learning based [10], and maximizing load balancing [11] algorithms.

Wardrop equilibrium algorithm models traffic steering problem as a distributed, non-cooperative, and dynamic load-balancing problem in the context of adversarial network equilibria. Based on the proper definition of latency functions representing the load of the access networks and consideration of constraints of the access network capabilities, the algorithm is proved, by Lyapunov arguments, to converge to an approximate Wardrop equilibrium, referred as the Backmann equilibrium in the literature, in which the latencies of the access networks are equalized. Simulation results validates the approach. (see [8] for the details)s

Reinforcement learning based algorithm models traffic steering and network selection problem as a markov decision process. It utilizes Q-learning based control design solution. We are currently working on the simulation of our proposed algorithm and the preliminary results validate the proposed solution. (see [9] for the details)

Maximization load balancing algorithm determines whether the gNB is overloaded or not by comparing it with a threshold based on the bandwidth usage ratio. Once a gNB is found overloaded, the algorithm sorts the users in the ascending order of reference signal received power (RSRP) and takes the first user from the list. Therefore, the algorithm triggers a traffic flow classification algorithm to classify delay tolerant and delay sensitive flows of the user. If delay tolerant flow is found, the proposed algorithm offloads the delay tolerant traffic to NTN satellite link and the satellite network delivers the data to that user. Then the algorithm again checks the bandwidth usage ratio to find whether the gNB is still overloaded or not. The above process continues while the gNB remains overloaded. Figure 8 shows the flow diagram of the proposed algorithm. (see [10] for the details)

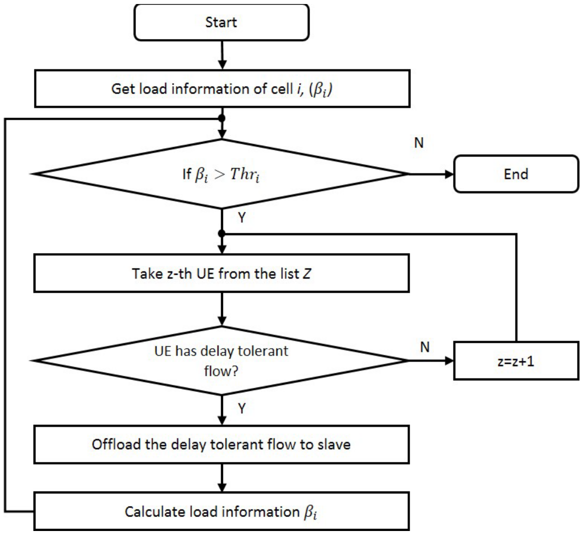

Start
Get load information of cell i, (β_i)
If $\beta_i > Thr_i$
N
End
Y
Take z-th UE from the list Z
UE has delay tolerant flow?
N
z=z+1
Y
Offload the delay tolerant flow to slave
Calculate load information β_i

Figure 8. Flow diagram for the maximizing load-balancing algorithm

V. CONCLUSION AND FUTURE WORK

We described 5G-ALLSTAR project research and development efforts on 5G multi-connectivity management and orchestration architecture and a proof-of-concept implementation. We provided initial R&D results. However, the project is still in the early stage of its R&D lifecycle and need more work in various aspects. We need to complete our PoC implementation and its functionality verification. Load-balancing algorithms simulations and performance verification have to be further improved. We are currently experimenting multiple algorithms but will eventually select an algorithm with best performance. Finally, we will test our PoC system in intercontinental R&D network between EU and Korea. For that we are currently finalizing the testing scenario. We will provide the detailed results of such enhancements in the future version of the paper

ACKNOWLEDGMENT

This work is supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2018-0-00175, 5G AgiLe and fLexible integration of SaTellite And cellulaR).

REFERENCES

[1] 3GPP TS 23.501, “3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; System Architecture for 5G System; Stage 2 (Release 15),” 2018-09.

[2] https://5g-allstar.eu

[3] S. Andreev *et al.*, “Intelligent access network selection in converged multi-radio heterogeneous networks,” *IEEE Wirel. Commun.*, 2014.

[4] 3GPP TR 28.808, “3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Management and Orchestration; Study on management and orchestration aspects with integrated satellite components in a 5G network (Release 16),” 2019-10.

[5] 3GPP TS 29.520, “3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; 5G System; Network Data Analytics Services; Stage 3 (Release 16),” 2019-09.

[6] 3GPP TR 28.809, “3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Management and Orchestration; Study on enhancement of Management Data Analytics(MDA) (Release 17),” 2019-10.

[7] 3GPP TS 28.530, “3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Concepts, use cases, and requirements (Release 15),” 2018-12..

[8] 3GPP TR 38.801, “3rd Generation Partnership Project; Technical Specification Group; Radio Access Network; Study on new radio access technology: Radio access architecture and interfaces (Release 14),” 2018.

[9] F. Delli Priscoli, et al., “Capacity-Constrained Wardrop Equilibria and Application to Multi-Connectivity in 5G Networks,” International Journal of Control, Submitted.

[10] A. Giuseppi, et al., “Traffic Steering and Network Selection in 5G Networks based on Reinforcement Learning,” European Control Conference 2020, Submitted.

[11] M. Mehedi Hasin, et al., “Load Balancing in 5G multi-RAT Networks by Offloading Delay Tolerant Flows,” KICS Summer Conference 2019, June 2019.