Dynamic Radio Access Selection and Slice Allocation for Differentiated Traffic Management on Future Mobile Networks

Claudia Carballo González, Ernesto Fontes Pupo, *Member, IEEE*, Luigi Atzori, *Senior Member, IEEE,* Maurizio Murroni, *Senior Member, IEEE*

doi: 10.1109/
TNSM.2022.3150978

Abstract— **The development of future wireless networks focuses on providing services with strict, dynamic, and diverse quality of service (QoS) requirements. In this sense, the network slicing paradigm arises as a critical piece on the efficient allocation and management of network resources, allowing for dividing the network into several logical networks with specific functionalities and performance. This paper aims at finding the best combination of access network and network slices over a heterogeneous environment to fulfill users' requests and optimize network resources usage. We propose the Dynamic radio Access selection and Slice Allocation (DASA) algorithm, flexibly adapted to network conditions, user priorities, and mobility behavior. DASA is based on a multi-attribute decision making (MADM) and analytical hierarchy process (AHP) to face the complex problem of network selection. Moreover, it uses a cooperative game theory approach to handle load balancing during overload situations. This work presents an integral solution that combines software-defined network (SDN) and network function virtualization (NFV) technologies to improve network performance and user satisfaction. DASA algorithm is evaluated through network-level simulations, focusing on flexibility and the effective utilization of network resources during network selection and load balancing mechanisms.**

Keywords— Heterogeneous wireless networks, NFV, Network slicing, QoS, SDN.

I. Introduction

According to the Cisco forecast [1], the number of connected devices will be more than three times the global population by 2023, and the machine-to-machine (M2M) connections will be half of the active connections. Moreover, in 2023 the global average broadband speed is expected to be more than double concerning the one in 2018.

Applications like 4K/8K ultra-high-definition video (VI), augmented and virtual reality (AR/VR), cloud gaming (GM), and internet of things (IoT) are characterized by different key performance indicators (KPI) [2, 3], imposing tight quality of service (QoS) requirements. Furthermore, heterogeneity is one of the keys of future wireless networks: current dynamic radio access and the "always best connected" (ABC) paradigm, between 5G and Wi-Fi 6 will be extended to non-terrestrial networks (NTNs) in 6G deployment [4, 5] making the management and exploitation of network resources even more complex.

In this cumbersome scenario, the network slicing paradigm emerges as a vital concept to introduce dynamism, isolation, and programmability over next-generation mobile networks [6, 7]. It partitions the single infrastructure into multiple logical network slices (NSs) where resources can be flexibly configured according to demand and heterogeneous network conditions.

On the other hand, the softwarization technologies software-defined network (SDN) and network function virtualization (NFV) complement the NS deployment [8, 9] to control and manage network resources in real-time efficiently. The SDN controller enforces intelligent decisions taking complete control of the network. At the same time, NFV guarantees to orchestrate computational and storage resources needed to instantiate network functions.

A user should be associated with one or more NSs via a macro-cell, small-cell, or Wi-Fi access point (AP) to satisfy his/her services requests. Additionally, a user has specific priorities according to his/her tariff plan and services preference. Therefore, the dynamic load balancing and the optimal selection of access network/NSs are challenging and necessary tasks to improve user satisfaction and optimize resource utilization.

Regarding the very complex issue raised above, recent works [10, 11] are focused on multi-attribute decision making (MADM) and analytical hierarchy process (AHP) to handle and reduce the complexity of the network selection problem [12]. However, these investigations do not consider overload scenarios neither users requesting multiple services simultaneously.

The topic of load balancing during an overload situation is also considered in past works [13, 14], using methods like the Markov decision process (MDP) [15] and the game theory approach [16], where they combine QoS parameters, device characteristics, and user profile. However, these works are focused on video service delivery without exploiting NS potentials.

Bearing the aforementioned methods as benchmarks, this paper proposes a novel Dynamic radio Access selection and Slice Allocation (DASA) algorithm over 5G heterogeneous networks and beyond. The selection process is based on MADM and AHP, considering network, user, device, and application profiles. Furthermore, DASA faces overload situations through a collaborative game theory approach, allowing slice reallocation on-demand and accepting more

Manuscript received on June 29, 2021. This work was supported in part by the Department of Electronical and Electronic Engineering (DIEE/UdR CNIT), University of Cagliari, 09123 Cagliari, Italy, and LACETEL, R&D Telecommunication Institute, 34515 Rancho Boyeros Ave., Boyeros, 19200, Havana, Cuba.

The authors are with the Department of Electrical and Electronic Engineering (DIEE/UdR CNIT), University of Cagliari, 09123 Cagliari, Italy (e-mails: claudia.carballogonz@unica.it, e.fontespupo@studenti.unica.it, l.atzori@diee.unica.it, murroni@diee.unica.it). E. Fontes is also with LACETEL, R&D Telecommunication Institute, Boyeros, 19200, Havana, Cuba (fontes@lacetel.cu).

© 20XX IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other

users with an adequate service perception. The algorithm is an integral solution to handle access network selection and slice allocation combining NS with SDN and NFV technologies.

The proposal is evaluated through network-level simulations using OMNeT++ [17] with the novel Simu5G [18] and Python tools. The validation is focused on DASA flexibility and the efficient utilization of network resources during the network selection method and load balancing strategy.

We prove that the algorithm is dynamically adapted to network conditions, user preferences, tariff plans, device resolution, and QoS constraints through the two deployed scenarios. Moreover, it allows for maximizing the number of users connected to the network with a QoS higher than or equal to the minimum requirements according to the user and application profiles during an overload situation.

The results show the improvement in applying two collaborative phases among access networks and among NSs. A throughput satisfaction balance is reached during the load balancing strategy, avoiding an abrupt degradation of user perception. The DASA validation demonstrates the differentiated treatment between premium and regular clients, ensuring a superior QoS for the users with high priority. Moreover, the algorithm achieves an efficient resource reallocation, affecting a minimum number of users at each saturation point.

The main contributions of this paper are summarized as follows:

- DASA integrates the network selection and load balancing mechanisms, managing an efficient slice allocation on heterogeneous wireless networks.
- The algorithm considers users requesting multiple services simultaneously and takes advantage of the network slicing paradigm by dynamically handling network conditions, user preferences, tariff plans, device resolution, and diverse QoS requirements.
- The load balancing process is based on game theory, applying two collaborative phases: among access networks and among NSs. This procedure guarantees an efficient slice reallocation on-demand, avoiding abrupt throughput degradations.

The remainder of this paper is structured as follows. Section II discusses related works in NS and softwarization approaches, network selection, and load balancing strategy. Section III introduces the proposed system and the DASA algorithm formulation. Section IV and V detail network selection and load balancing mechanisms, respectively. Section VI describes simulation and deployment scenarios, whereas section VII discusses the simulations results. Finally, in section VIII, the document is concluded.

II. Related Work

This section surveys the state of the art related to NS and softwarization approaches, network selection, and load balancing strategy.

A. NS and Softwarization Approaches

The research community stands out the integration of SDN, NFV, and network slicing paradigms to introduce flexibility, programmability, and dynamism in 5G networks and beyond. This subsection covers some of the most recent works related to NS and softwarization technologies.

In [19], the authors highlighted the value of networking automation and intelligence to dynamically meet tight service requirements over future networks. This work is oriented to improve network management and operations efficiency through softwarization technologies, but without considering the potential of NS. Instead, Barakabitze *et al.* [8] reviewed the 5G NS academic and industry solutions using SDN and NFV technologies and the standardization efforts. They presented a deep state of the art analysis, highlighting architectural approaches in practical implementations and deployment strategies.

Taleb *et al.* [20] proposed an NS management and orchestration architecture incorporating SDN and NFV components to the 3rd Generation Partnership Project (3GPP) NS management. In [21], the authors proposed a QoS framework for NS based on SDN and NFV technologies. Their proposal is based on monitoring network status to compute a path and allocate QoS resources. Yousaf *et al.* [22] presented a mobility management extension for NFV orchestration to adapt the NS capacity on-demand and select the suitable template for slice instantiation and the correct mobility scheme according to the users' mobility behavior. Ahmed *et al.* [23] introduced an architecture framework that combines SDN, NFV, and satellite NS to distribute network resources on-demand. In addition, they defined several services that must be attended to considering available capacity and QoS requirements. In [24], the authors presented a survey about NS modellings in the radio access network (RAN), core network, and transport network domains. They explained the end-to-end (E2E) NS process from service requests to NS deployment and life-cycle management.

The papers mentioned above represent a good benchmark to understand the integration of our proposal into the different state of the art architectures. Furthermore, these works distinguish the high value of merging NS and softwarization technologies over next-generation wireless networks.

B. Network Selection Mechanism

In a heterogeneous scenario, due to the limited nature of resources and strict services requirements, selecting the best access networks and NSs to satisfy the user demand has been in the crosshairs of many researchers. This subsection presents some of the most recent works dealing with the network selection mechanism with particular attention to the MADM method.

In [10], Desogus *et al.* presented the TYDER algorithm, a network selection method based on MADM and AHP. They considered network reputation in the context of diverse traffic requirements, and they analyzed four types of applications: VI, GM, navigation, and IoT. Montalban *et al.* [11] proposed a similar approach based on MADM for the network selection, focused on a convergent architecture for broadcast, broadband, and cellular services. This proposal considered throughput, delay, packet success rate, and spectral efficiency parameters. In [25], the authors presented a multi-criteria analysis for IoT applications. They also used MADM and AHP methods, combining parameters such as cost, number of customers, privacy, and availability.

The above investigations represent an excellent approach to the network selection method. However, they did not address the users' requests for multiple services simultaneously, and they did not include the NS paradigm. Instead, in [26], the authors presented a multi-criteria analysis to select the best NS among candidate slice instances. Furthermore, the proposal had in consideration memory and CPU resources, as well as throughput and delay.

Alternatively, Bakmaz *et al.* [27] proposed a Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) to select the best NS among available slices. TOPSIS considers the following steps: normalization, weighting calculation, and alternative ranking [28]. In [29], the authors proposed a genetic algorithm (GA) [30] to select the optimal base station (BS) and NS. The proposal is validated by comparing with the traditional received signal strength (RSS) criterion [31] and the greedy algorithm [32]. However, they considered that a slice in the network could be only associated with one BS for proving the proposal. Moreover, they assumed that a user only requests one service flow, limiting the NS advantages.

References [26, 27, 29] represent a good approach to NS's access and handover process. Nevertheless, they did not inquire into QoS metrics' different behavior for several possible applications in the selection process.

On the other hand, Sun *et al.* [33] proposed a reinforcement learning algorithm to drive the handover process for RAN slicing. They defined the coverage area for each NS, differing in terms of data rate and delay. The proposal is compared with the other three baselines' methods: maximum signal-to-interference-plus-noise ratio (Max-SINR), which represents a variation of the traditional RSS-based handover scheme, NS-prioritization, and a modified version of their proposed method.

C. Load Balancing Strategy

Load balancing is a crucial strategy to optimize resource utilization and to improve users' perception. Therefore, it plays a critical role, especially in overload situations, where the resources need reallocation to accept new users and do not degrade the QoS below a certain threshold.

In [34], the authors focused on the principles and models of load balancing in 5G NS. They categorized different mathematical models of resource allocation, dividing them into game-theoretic economic models [16], prediction models [35-37], and robustness and failure recovery models [38, 39].

Specifically, the game theory approach has been extensively used in networking research problems, and it is suitable to apply in allocation problems over 5G heterogeneous networks and beyond [16, 40, 41]. This mathematical framework is a powerful tool to model players' interaction, attempting to make decisions that maximize each entity's utility. In particular, Han *et al.* [42] described the potential of cooperative game theory to handle a wide range of applications in wireless communication systems. Additionally, one of the main branches of the cooperative game is the coalition theory, deeply described in [43].

Desogus *et al.* [14] proposed a cooperative game theory solution for load balancing. Moreover, they presented a heterogeneous network selection algorithm based on MADM, considering the user, device, application, and network profiles. In [44], the authors presented a game theory method to face the NS admission and overload control. The proposal is based on dynamic share allocation, and it is compared with a fixed share alternative.

On the other hand, Anedda *et al.* [13] presented an MDP solution to perform resource allocation. They considered diverse users' priorities (business and typical) and different screen resolutions. The overload situation is handled by gradually decreasing the throughput of current users to accept new customers on the network.

The papers [13, 14] focused on video content delivery, and they did not consider the network slicing paradigm. Moreover, unlike reference [13], [14] and [44] did not include a user priority differentiation.

Compared to the previous works, DASA is an integral solution that guarantees the dynamic and efficient resource allocation over 5G heterogeneous networks and beyond, handling the network selection method and the load balancing strategy. DASA considers users with different priorities requesting multiple services simultaneously. Moreover, the proposal takes advantage of NS, SDN, and NFV potentials, allowing an adequate QoS and management of critical services.

III. Overview of the Proposed System

In the following subsections, we illustrate the considered scenario, we define the used notation, and we provide an overview of the proposed algorithm.

A. Considered Scenario

The proposed algorithm considers an E2E slices deployment based on the softwarization technologies SDN and NFV, with particular application to the RAN domain. The proposed solution is set on the reference architecture [45, 46] shown in Fig. 1, which includes the RAN with several BSs, the transport network layer, the mobile packet core network, and the cloud computing infrastructure. The architecture considers different service types (i.e., Services 1-L in Fig. 1) demanding specific resources and attributes, and it provides an E2E connection service according to the users' requests.

In this softwarized system, DASA is integrated into the NS Management and Orchestration module, interacting with the Slice Controller during the network selection and load balancing processes. The Slice Controller has complete control of the network. It oversees the coordination of the services requests and drives resource allocation.

The DASA algorithm is triggered when the Slice Controller detects a new user's services request (e.g., connecting a group of IoT devices, streaming 360-degree video), when a user wants to change the service type and requirements, or when it is necessary a handover process or a resource reallocation due to the degradation of the current network's QoS below a

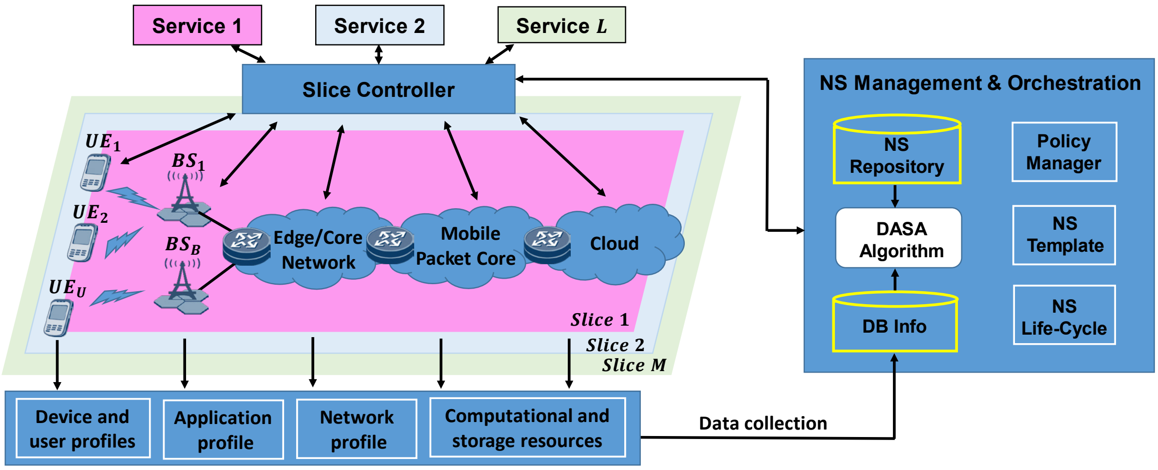

Fig. 1. NS architecture with the DASA algorithm.

certain threshold. These thresholds are imposed in terms of throughput, delay, jitter, packet loss ratio, and energy consumption, depending on application and user profiles.

The objective of the DASA module is to select the best combination of access network (i.e., which BS to be associated) and NSs to satisfy the user's requests and handle a load balancing strategy during an overload situation. To determine the best combination to satisfy the user's requests and maximize the resulting QoS, DASA requires access to the Database Information (DB Info) and the NS Repository. The former gathers in real-time the following data:

- **User profile:** contains the users' preferences and their tariff plans.
- **Device profile:** specifies the device characteristics, such as the screen resolution, current position, and mobility behavior.
- **Network profile:** collects information about ID, type, and QoS of the available wireless networks.
- **Application profile:** details the services/NSs performance that users are currently using.

It is important to highlight that, following the General Data Protection Regulation (GDPR) [47], all the collected data must be kept anonymous to protect users' privacy. Specifically, an identification number is associated with the user terminal which is not linked to any information that may identify the user. Additionally, this data is collected once user consent is given.

The NS Repository contains information about the active NSs in each architecture layer, their supported service instances, and the allocated virtual and physical resources. This information is used to check if the slice to accommodate a requested service exists.

DASA sends the results to the NS Life-Cycle, which manages the topology/resources update, the creation, and deletion of NSs. The NS creation is based on the NS Template, which stores the attributes that characterize the topology of NS services. Additionally, the Policy Manager oversees the NS policies enforcement, such as the access rules and the creation and expiration times.

B. *Algorithm Notation*

We assume that the network serves a set of U users denoted by $\mathbb{U}$, with the sub-index $u \in \{1, 2, \dots, U\}$. The users are randomly distributed in two priority levels according to their tariff plan. Priority "high" corresponds to premium clients, which pay more to guarantee the best possible QoS level. In contrast, priority "low" corresponds to regular users who pay less and are satisfied with the minimum quality levels.

The set of B BSs defined by $\mathbb{B}$, with the sub-index $b \in \{1, 2, \dots, B\}$, corresponds to the available RANs, which could be small-cells, macro-cells, or Wi-Fi APs. The set of M NSs denoted by $\mathbb{M}$, with the sub-index $m \in \{1, 2, \dots, M\}$, can be available or not in different BSs. They can host one or more service instances, defined in terms of QoS parameters and specific demanding resources (e.g., ultra-low latency or high-data rate demanding applications). The set of NSs to accommodate the user equipment (UE_u) requests is denoted by $\mathbb{M}_u$, where $\mathbb{M}_u \subseteq \mathbb{M}$.

We assume a set of L possible services defined by $\mathbb{L}$, with the sub-index $l \in \{1, 2, \dots, L\}$. Besides, the set of services requested by the user UE_u is defined by $\mathbb{L}_u$, where $\mathbb{L}_u \subseteq \mathbb{L}$. In this work, we consider four services: VI, AR, IoT, and GM, characterized by specific functionalities and resources. Therefore, a user can request from one to four of these services simultaneously. We assume that each service is associated with one NS (i.e., one NS for each of the four typologies per user).

According to the services' preferences, the most important service for the user UE_u is classified as primary service (PS), and the rest as secondary services (SSs). This preference relation is represented by the sub-index o, where $o = 1$ identifies the PS, whereas $o > 1$ represents the SSs in order of preference.

The $p^u_{m,o}$ is the priority value of the requested NS_m by the user UE_u. The highest priority slice for the user hosts his/her PS, and we assign to this slice the highest weight. At the same time, the priority of the secondary services is weighted in decreasing order according to the user petition (i.e., SS_1 priority is higher than SS_2). The 3GPP defines that the UE does not need to support more than eight slices simultaneously [48]. Therefore, we assume the priority weights of the users' requested services according to Table I. The $p^u_{m,o}$ follows the rule expressed in

$$\sum_{m\in\mathbb{M}_u} p^u_{m,o} = 1. \quad (1)$$

To evaluate the network conditions to satisfy each requested service (with preference o configured in the NS_m), we define the individual score $S_{b,m,o}$. The score for each requested service is represented by utility functions (UF) and their respective weights (w). The UFs represent the most sensitive attributes that characterize the QoS (i.e., throughput, delay, jitter, packet loss ratio, and energy consumption). Additionally, the weights denote the importance of such QoS parameters. The individual score is defined as follows

$$S_{b,m,o} = \sum_{uf\in\mathbb{UF}} w_{m,uf} * UF_{m,uf}, \quad (2)$$

with

$$\sum_{uf\in\mathbb{UF}} w_{m,uf} = 1, \quad (3)$$

where uf is the sub-index to iterate through the set $\mathbb{UF}$ and their corresponding weights. The cardinality of $\mathbb{UF}$ is denoted by n, representing the number of sensitive attributes associated with each service.

C_b denotes the total capacity in the BS_b. Moreover, the total capacity used by the NS_m in the BS_b is the sum of the assigned throughput ($Th^u_{b,m}$) for each user UE_u, denoted by

$$C_{b,m} = \sum_{u\in\mathbb{U}} Th^u_{b,m}. \quad (4)$$

As a constraint, the used capacity for the NSs configured in the BS_b must be less than or equal to the total capacity of the BS_b

$$\sum_{m\in\mathbb{M}} C_{b,m} \leq C_b. \quad (5)$$

The available capacity in the BS_b is Ca_b, and it is represented by

$$Ca_b = C_b - \sum_{m\in\mathbb{M}} C_{b,m}. \quad (6)$$

Additionally, the maximum bit rate supported by the user UE_u accessing the access network BS_b is denoted by $bit\_rate^u_b$.

The availability of resources for the NS_m in the BS_b is denoted by $r_{b,m,o} \in \{0,1\}$, where 0 means not enough resources to support the NS_m. The accessibility of the NS_m via the BS_b is denoted by $d_{b,m,o} \in \{0,1\}$, where 0 means that the NS_m is not available from BS_b, due, for example, to missing functionality or to the impossibility to access the service.

TABLE I.
PRIORITY WEIGHTS FOR NS USAGE

Usage	PS	SS_1	SS_2	SS_3	SS_4	SS_5	SS_6	SS_7
1	8/8							
2	6/8	2/8						
3	6/8	2/12	1/12					
4	6/8	3/24	2/24	1/24				
5	4/8	4/20	3/20	2/20	1/20			
6	4/8	5/30	4/30	3/30	2/30	1/30		
7	4/8	6/42	5/42	4/42	3/42	2/42	1/42	
8	4/8	7/56	6/56	5/56	4/56	3/56	2/56	1/56

From the previous explanation, we define the possible states of NS_m in the BS_b as $St_{b,m} \in \{0, 1, 2\}$, where the meaning of these values is displayed in Table II. The NS_m is activated in the BS_b if it is accessible and is being used. This information is contained in the NS Repository.

As a result of the algorithm, the throughput satisfaction (Th_{sat}) for the user UE_u request in the BS_b can be expressed by

$$Th_{sat} = \sum_{m\in\mathbb{M}_u} \frac{Th^u_{b,m}}{Thmax^u_m} \quad (7)$$

where $Thmax^u_m$ is the maximum throughput supported by each requested service. Additionally, the satisfaction degree (SD) is the ratio between the number of assigned slices and the total requested services by the user UE_u.

For better understanding, the mathematical notations used in the paper are summarized in Table III.

C. Overview of the Proposed Algorithm

DASA selects the best combination of access network and NSs over a heterogeneous environment to fulfill users' requests and optimize network resources usage. Fig. 2 shows the DASA algorithm flowchart, highlighting the most critical processes detailed in the following sections IV and V.

DASA is triggered when the Slice Controller detects a new user request or that a degradation on QoS parameters reaches

TABLE II.
POSSIBLE NS STATES

$St_{b,m}$	Accessible	Activated
0	✗	✗
1	✓	✗
2	✓	✓

"✓" the condition is met.
"✗" the condition is not met.

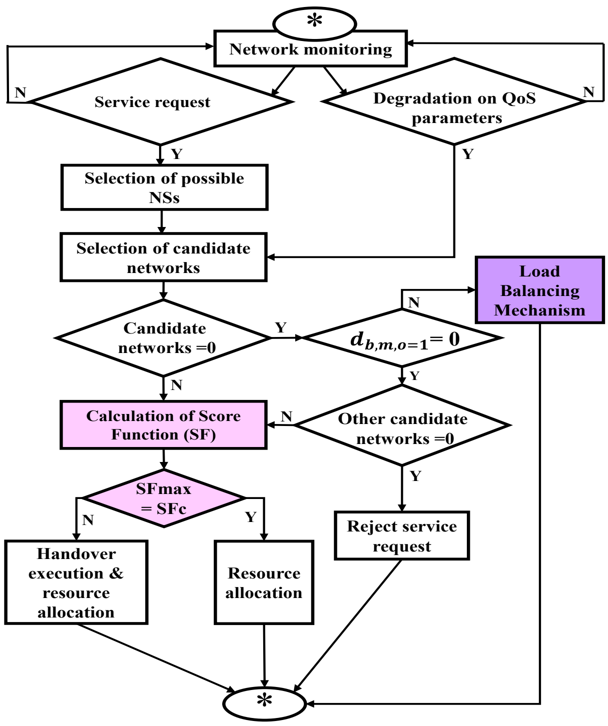

Fig. 2. The DASA algorithm flowchart.

TABLE III.
MATHEMATICAL NOTATIONS

Notation	Definition
$\mathbb{U}(u \in \{1,2,\ldots,U\})$	Set of U users, where the user u is denoted by UE_u
$\mathbb{B}(b \in \{1,2,\ldots,B\})$	Set of B BSs
$\mathbb{M}(m \in \{1,2,\ldots,M\})$	Set of M NSs
$\mathbb{L}(l \in \{1,2,\ldots,L\})$	Set of L services
$St_{b,m}$	Possible states of NS_m in the BS_b
$C_{b,m}$	Capacity used by the NS_m in the BS_b
Ca_b	Available capacity in the BS_b
$bit\_rate_b^u$	Maximum bit rate supported by UE_u accessing to BS_b
$r_{b,m,o}$	Availability of resources for the NS_m with preference o in the BS_b
$d_{b,m,o}$	Accessibility of the NS_m with preference o via the BS_b
CAN/CAN^*	Candidate access network / candidate access network during overload situation
$p_{m,o}^u$	Priority of the NS_m with preference o requested by the UE_u
$SF_{b,u}$	Score function of BS_b for the user UE_u service request
$S_{b,m,o}$	Score for a service with preference o configured in the NS_m via the BS_b
$\mathbb{UF}(uf \in \{1,2,\ldots,UF\})$	Set of UF utility functions
$w_{m,uf}$	Weight for each UF in the analyzed service
α, β	Tuned steepness parameters in the UF
$Th_{b,m,o}^u$	Throughput assigned to UE_u for the service with preference o allocated in NS_m via the BS_b
$D_{b,m,u}$	Delay from UE_u to access the requested service allocated in the NS_m via the BS_b
$PLR_{b,m,u}$	Packet loss ratio for the NS_m via the BS_b
$J_{b,m,u}$	Jitter from UE_u to access the requested service allocated in the NS_m via the BS_b
$Ec_{b,m,u}$	Energy consumption by UE_u using the service allocated in NS_m via the BS_b
Th_{sat}	Throughput satisfaction
SD	Satisfaction degree
A/A_{norm}	Pairwise comparison' s matrix/ Normalization matrix
$\mathbb{S}(s \in \{1,2,\ldots,S\})$	Set of S coalitions
$v(S_s)$	Characteristic function of S_s coalition
S_{GC}	Grand coalition structure
$P_{Th_b}^N$	Normalized value of the potential resources that BS_b can release
Θ	Fraction of the supported services request between the total requested services
$\lambda_{t,t+1}$	Actions' sequences available to perform on a current UE_u
Ƒ	Minimum resources that NSs need to free up in order to satisfy the new service request
Th_r, P_r	Released resources, total released resources
ɐ	Residual value
$Uaff_N$	Normalized value of affected users
ρ	Rule of order

a certain threshold according to service performance and user's priority. According to Fig. 2, if there is a new service request, first, the algorithm identifies the NSs able to accommodate the petition. Then it selects the candidate access networks ($CANs$) among the access networks in the user's coverage area (i.e., small-cell, macro-cell, or Wi-Fi AP).

One access network BS_b is considered a candidate network if it supports at least the NS corresponding to the user's PS. Moreover, the $bit\_rate_b^u$ and the Ca_b must be higher than or equal to the minimum required throughput for the PS. Additionally, the delay, jitter, packet loss ratio, and energy consumption of the user accessing the PS via BS_b must be less than or equal to the maximum admissible values.

Once the $CANs$ have been identified, DASA determines the best access network to fulfill the user request. The algorithm computes a score function (SF) for each CAN, and the BS_b with the highest value (SF_{max}) is selected. If the SF_{max} value does not belong to the user's current network (SF_c), the handover process will be executed.

Let us suppose there are no $CANs$ due to the NS's inaccessibility for the primary service ($d_{b,m,o=1}$= 0). In that case, DASA selects other candidate networks (CAN_{others}) that support the secondary services, and it applies the rest of the algorithm according to Fig. 2. If there are no $CANs$ due to a lack of resources for the PS ($r_{b,m,o=1}$ = 0), the load balancing mechanism is used, which is presented in section V.

When the network has enough capacity, DASA assigns the maximum resources and does not consider the user's priority. On the other hand, during a load balancing mechanism, the algorithm applies different strategies to adjust the throughput from the current value to the imposed threshold according to the service performance and the user's tariff plan. This resource allocation strategy is applied according to the defined service level agreement (SLA) among provider and tenants.

DASA never assigns more resources than the maximum required by the service ($Thmax_m$). The algorithm ensures the maximum possible $Th_{b,m}^u$ for the user UE_u according to the Ca_b, $bit\_rate_b^u$, service requirement, device resolution, and tariff plan.

IV. NETWORK SELECTION

As highlighted in Fig. 2, the most important element of the network selection process is the calculation of SF for each CAN to select the most suitable combination of access network and NSs. This process is detailed in the following subsection.

A. Calculation of Score Function

The SF calculation applies MADM to combine multiple normalized attributes considering the NSs availability, network conditions, user and application profiles. As we assumed in subsection III-B, an NS only allocates one service. Therefore, the SF is defined by

$$SF_{b,u} = \sum_{m \in \mathbb{M}_u} r_{b,m,o} * d_{b,m,o} * p_{m,o}^u * S_{b,m,o} \quad (8)$$

where the notations are presented in subsection III-B. As we previously described, the individual score $S_{b,m,o}$ for each requested service is represented by utility functions and their respective weights. In this work, we consider the utility functions (UF_{up} and UF_{down}) based on sigmoid functions [49-51], expressed by

$$UF_{up}(x) = \begin{cases} 0, \ x < x_{min} \\ 1 - \left(\dfrac{\left(\frac{x_{max}-x}{x_{max}-x_{min}}\right)^{\alpha}}{1+\left(\frac{x-\beta * x_{min}}{x_{max}-x_{min}}\right)^{\alpha}} \right), \ x_{min} \le x \le x_{max} \\ 1, \ x > x_{max} \end{cases} \quad (9)$$

$$UF_{down}(x) = \begin{cases} 0, & x > x_{max} \\ \frac{\left(\frac{x_{max}-x}{x_{max}-x_{min}}\right)^{\alpha}}{1+\left(\frac{x-\beta * x_{min}}{x_{max}-x_{min}}\right)^{\alpha}}, & x_{min} < x \leq x_{max} \\ 1, & 0 \leq x \leq x_{min} \end{cases} \quad (10)$$

where (9) represents the upward criteria and describes the throughput (Th). On the contrary, (10) denotes the downward criteria and describes the delay (D), jitter (J), packet loss ratio (PLR), and energy consumption (Ec). The variables x_{min} and x_{max} are the minimum and maximum values defined for each service. The functions are dimensionless and range from 0 to 1. The variables α and β are tuned steepness parameters depending on the application type. According to [49], we assume that $\alpha \geq 2$ and $\beta > 0$.

The delay from the user UE_u to access the requested service is expressed by

$$D^u_{b,m} = Duser_{u,b} + Dservice_{b,m} \quad (11)$$

where $Duser_{u,b}$ is the delay from user to BS, and $Dservice_{b,m}$ is the delay from BS to the server where the service is allocated.

The packet loss ratio is defined by

$$PLR^u_{b,m} = \frac{Missed_{packets}{}^u_{b,m}}{Total_{packets}{}^u_{b,m}} \quad (12)$$

where $Missed_{packets}$ is the number of lost packets from the total number of transmitted packets ($Total_{packets}$).

To obtain the jitter, we apply

$$J^u_{b,m} = Dpeak^u_{b,m} + 2 * F * Rrms^u_{b,m} \quad (13)$$

where $Dpeak$ is the deterministic jitter according to systematic effects (e.g., inter-symbol interference and periodic jitter). $Rrms$ is the random jitter resulting from the accumulation of random processes, and it requires statistical analysis to be quantified. Additionally, F is a factor tabulated according to different bit error rate (BER) tolerances [52].

The energy consumption, expressed in Joule (J), is derived as

$$Ec^u_{b,m} = t * \left(r_t + Th^u_{b,m} * r_d\right) \quad (14)$$

where t is the transaction time in seconds, obtained from the duration of the application; r_t is the mobile device's energy consumption per unit of time in Watt (W); $Th^u_{b,m}$ is the assigned Th for the user UE_u to access the NS_m in the BS_b; and r_d is the energy consumption rate for the received stream expressed in J/Kbyte [10].

Fig. 3 shows the analyzed services and the sensitive attributes that affect in different proportions their QoS. We consider n = 3 for IoT and n = 5 for VI, AR, and GM, where n is the number of comparable utility functions in the set $\mathbb{UF}$ as previously defined.

In the case of the analyzed multimedia services, Th is the attribute with the most significant impact. According to the IoT scenario, the applications can be classified into machine or human-oriented, delay-sensitive or not, and with different Th requirements [53]. In this proposal, as shown in Fig. 3, we assume IoT applications characterized by a massive number of connected devices transmitting a low data rate and non-delay sensitive data (e.g., smart grid, smart home, and smart city applications).

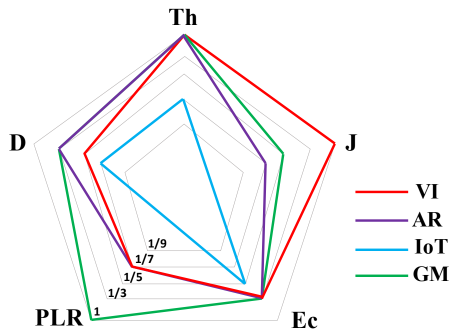

Fig. 3. UFs for VI, AR, IoT, and GM.

To define each service's weights (presented in (3)), we apply the AHP and the Saaty-scale [54], used in many scenarios of multi-criteria decisions. AHP allows decomposing a complex problem into a hierarchical structure easier to solve. The Saaty-scale permits to establish priorities among pairs according to their relevance. Table IV shows the possible assigned values according to [54].

The pairwise comparison's matrix is defined as

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} \quad (15)$$

where $a_{i,j}$ represents the importance criterion among different UFs according to the Saaty-scale values.

Table V details the pairwise comparison's matrix for AR. Th is the most important factor (value 1). In contrast, D and Ec are a little less critical (value 1/3 with respect to Th), and PLR and J are less important (value 1/7 concerning Th).

Later it is necessary to obtain the normalization matrix and finally, the weight for each UF, defined as

$$A_{norm} = \begin{bmatrix} \frac{a_{1,1}}{\sum a_{i,1}} & \cdots & \frac{a_{1,n}}{\sum a_{i,n}} \\ \vdots & \ddots & \vdots \\ \frac{a_{n,1}}{\sum a_{i,1}} & \cdots & \frac{a_{n,n}}{\sum a_{i,n}} \end{bmatrix}, \quad (16)$$

TABLE IV.
SAATY-SCALE VALUES

Scale	Definition
1	Equal importance
3	Moderate importance of one over another attribute
5	Strong importance
7	Very strong importance
9	Extreme importance
1/3	Little less critical one over another
1/5	Fairly less importance
1/7	Definitely less importance
1/9	Extremely less importance

TABLE V.
PAIRWISE COMPARISON'S MATRIX FOR AR

	Th	D	Ec	J	PLR
Th	1	3	3	7	7
D	1/3	1	1	5	5
Ec	1/3	1	1	5	5
J	1/7	1/5	1/5	1	1
PLR	1/7	1/5	1/5	1	1

$$w_i = \frac{\sum_{j=1}^{n} a_{i,j}}{n}. \quad (17)$$

Table VI summarizes the final obtained weights for each analyzed service and its fundamental factors for the QoS through the AHP process.

The SF process calculation is described in Algorithm 1, dynamically adapted to user preferences and network conditions.

V. Load Balancing Mechanism

If there are no candidate networks due to limited resources, the load balancing mechanism is applied. Our proposal is based on the cooperative game theory to benefit more users through dynamic resource allocation. We define the coalition set as $\mathbb{S} = \{S_1, S_2, \dots, S_s\}$, where S_s is a subset of cooperative players, and $v(S_s)$ is the characteristic function that describes the payoff associated with the coalition S_s.

The game initiation trigger is the reach of a saturation point in the network and if a new service request is received. This mechanism runs in two consecutive phases:

1- Collaboration among $CANs$ (named $CANs^*$ during the load balancing process), based on a canonical game [42], to determine the most suitable BS to satisfy the new service request.

2- Collaboration among NSs on the selected BS, based on a dynamic coalition formation game [42], to free up resources according to the required throughput and the conditions of current users.

The overall utility function is to maximize the number of users connected to the network with a QoS higher than or equal to the minimum requirements according to their tariff plans. Fig. 4 shows the load balancing flowchart, whose details are presented in the following subsections V-A and B.

TABLE VI.
UFs Weights for VI, AR, IoT and GM

Weight/ Service	w_{Th}	w_D	w_J	w_{PLR}	w_{Ec}	$\sum_{uf\in\mathbb{UF}} w_{m,uf}$
VI	0.36	0.08	0.36	0.04	0.16	1
AR	0.47	0.21	0.055	0.055	0.21	1
IoT	0.2	0.2	-	-	0.6	1
GM	0.34	0.13	0.06	0.34	0.13	1

Algorithm 1: Score Function Calculation

Result: Return the best candidate network (SF_{max}).
Input: ServiceType = List of services request;
CandidateNetworks = List of candidate networks;
begin
  SF_{max}= 0;
  $SF_{b,u}$ = 0;
  for each b in CandidateNetworks **do**
    for each m in ServiceType **do**
      if ServiceType == VI **then**
        $S_{b,m,o} = 0.36 * (UF_{Th} + UF_J) + 0.16 * UF_{Ec} + 0.08 * UF_D + 0.04 * UF_{PLR}$
      end
      if ServiceType == AR **then**
        $S_{b,m,o} = 0.47 * UF_{Th} + 0.21 * (UF_D + UF_{Ec}) + 0.055 * (UF_J + UF_{PLR})$
      end
      if ServiceType == IoT **then**
        $S_{b,m,o} = 0.6 * UF_{Ec} + 0.2 * (UF_{Th} + UF_D)$
      end
      if ServiceType == GM **then**
        $S_{b,m,o} = 0.34 * (UF_{PLR} + UF_{Th}) + 0.13 * (UF_D + UF_{Ec}) + 0.06 * UF_J$
      end
    end
    $SF_{b,u} = \sum_{m\in\mathbb{M}_u} r_{b,m,o} * d_{b,m,o} * p_{m,o}^{u} * S_{b,m,o}$
    if $SF_{b,u} > SF_{max}$ **then**
      $SF_{max} = SF_{b,u}$
    end
  end
end

A. Collaboration among Candidate Networks

When the network reaches the overload situation and a new user has a service request, the DASA algorithm selects the $CANs^*$ to determine the best BS to free up resources and attend the petition. The collaboration among the possible BSs (players) represents a canonical coalition game or grand coalition (GC) structure. It assumes that forming a larger coalition (S_{GC}) among the $CANs^*$ cannot be worse than acting alone. Therefore, this formation is characterized by the superadditive property, which means that the payoff $v(S_{GC})$ is always higher than or equal to the payoff value of any disjoint set of coalitions.

The $CANs^*$ are the access networks that meet the requirements of the $CANs$, except that they do not have enough available resources to satisfy the petition. Moreover, the $CANs^*$ must comply

$$P_{Th_b} \geq Thmin_{m,o=1}^{u} \quad (18)$$

where P_{Th_b} represents the potential resources that BS_b can release until all the users belonging to it have the minimum possible Th.

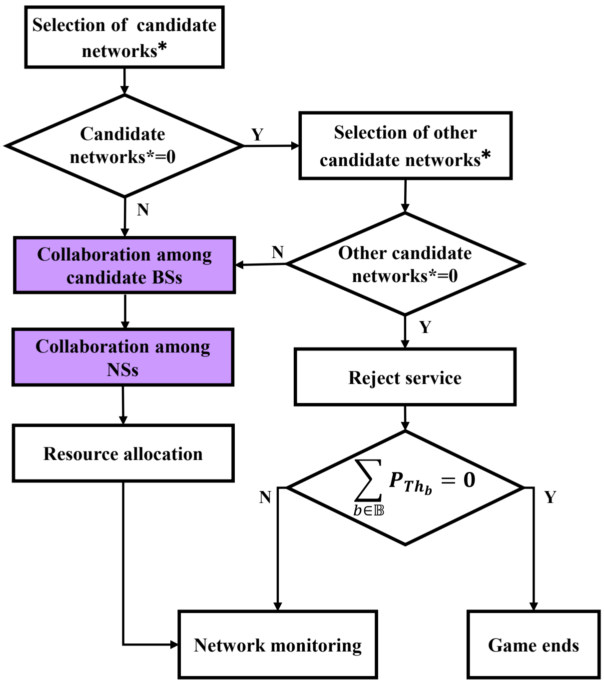

Fig. 4. The load balancing mechanism flowchart.

The normalization of P_{Th_b} ($P_{Th_b}^N$) is expressed by

$$P_{Th_b}^N = \begin{cases} 0, \; P_{Th_b} < Thmin_{m,o=1}^u \\ 1, \; P_{Th_b} \geq \sum_{m \in \mathbb{M}_u} Thmax_m^u \\ \dfrac{P_{Th_b}}{\sum_{m \in \mathbb{M}_u} Thmax_m^u}, \text{otherwise.} \end{cases} \quad (19)$$

The $CANs^*$ are evaluated according to the resulting fraction of the supported services request between the total requested services (Θ), $P_{Th_b}^N$ and Th_{sat} values to select the access network that could better satisfy the new user petition.

According to the user priority and similar users in the network, the algorithm follows a fairness rule among the $CANs^*$ looking to satisfy the new request with the best QoS possible. Moreover, the BSs play a collaborative attitude trying to balance their resources, avoiding that Th_{sat} abruptly decreases in one access network regarding the others. Therefore, it guarantees a stable and fair grand coalition, where the players do not have incentives to leave this structure.

According to the above assumptions, the utility score function (SF_b^*) for each CAN^* is

$$SF_b^* = 0.2 * (\Theta + Th_{sat}) + 0.6 * P_{Th_b}^N. \quad (20)$$

Equation (20) is determined by MADM and AHP combination, where $P_{Th_b}^N$, Θ and Th_{sat} are the utility functions and range from 0 to 1. In this case, we assume that $P_{Th_b}^N$ is a little more critical for the decision criteria concerning Θ and Th_{sat} (Table VII). Therefore, the weights (i.e., 0.2 and 0.6) represent the relevance values attributed to each associated parameter resulting from (15), (16), and (17).

The CAN^* with the SF_{max}^* is the BS with the best conditions to satisfy the services request. Therefore, SF_b^* quantifies the preferences between different outcomes [55]. One BS has strong preference respect to others (e.g., $BS_1 \succ BS_2$) if $SF_b^*(BS_1) > SF_b^*(BS_2)$.

If there are not $CANs^*$ due to insufficient resources to release for the PS or $d_{b,m,o=1}$= 0, DASA selects the CAN_{others}^* that support the SSs. In this case, the CAN_{others}^* are the players who cooperate to determine the best BS to serve the new user petition. P_{Th_b} must be higher than or equal to the minimum Th required from one of the SSs.

The above collaboration among BSs is described in Algorithm 2.

Algorithm 2: Collaboration among BSs

Result: Return the best candidate access network to attend the new service request.
Input: ServiceType = List of services request;
CandidateNetworks* = List of CAN^* or CAN_{others}^*;
begin
  SF_{max}^*= 0;
  SF_b^*= 0;
  for each b **in** CandidateNetworks* **do**
    SF_b^*= $0.2 * (\Theta + Th_{sat}) + 0.6 * P_{Th_b}^N$
    if $SF_b^* > SF_{max}^*$ **then**
      $SF_{max}^* = SF_b^*$
    end
  end
end

B. Collaboration among NSs

This phase is based on a dynamic coalition formation according to [42]. This kind of cooperation is an efficient strategy to optimize the network's social welfare. It is subject to environmental changes such as the number of players, their characteristics, and network conditions.

The main elements of this phase of the game are:

- Set of players, which are the available NSs in the selected BS, and the action space is the combination of all resources (services) of users belonging to this access network.
- The coalition set $\mathbb{S} = \{S_1, S_2, \dots, S_s\}$, where S_s is the subset of players that cooperate in reallocating slice resources and accepting new users in the network. The coalition must guarantee the throughput of the served users to be higher than or equal to a certain threshold defined by their priorities.
- Actions' sequences $\lambda_{t,t+1}$ available to perform on a current user UE_u. These actions represent the transition strategy from the assigned throughput to the minimum throughput possible.

VI, AR, and GM are considered high-demanding data rates multimedia services. Therefore, the actions consider the transition through the states $t = 4K, 2K, FHD, HD$. Table VIII shows the data rates constraints for each resolution according to [13, 56-59]. On the other hand, we consider IoT applications with low data rate demand. According to [53], we assume the states t = 1 Mbps, 500 kbps, 100 kbps.

The presented sequential actions allow to gradually decrease the data rate without abruptly affecting users' QoS. Moreover, a user whose throughput has been affected due to an overload situation can revert to the previous state when new resources are available on the network.

- Minimum cost Ƒ that represents the minimum resources (decision point) that NSs need to free up in order to satisfy the new services request and to guarantee a minor impact on the current users' satisfaction. Therefore, the coalition is considered only if the sum of released resources (Th_r) of its members is higher than or equal to the required throughput by a new services request. Ƒ depends on the new user's tariff plan, network conditions, and current data rate of similar users.

To affect as little as possible the users' perception, the

TABLE VII.
PAIRWISE COMPARISON'S MATRIX

	$P_{Th_b}^N$	Θ	Th_{sat}
$P_{Th_b}^N$	1	3	3
Θ	1/3	1	1
Th_{sat}	1/3	1	1

TABLE VIII.
APPLICATIONS DATA RATES

	VI (Mbps)		AR (Mbps)		GM (Mbps)	
	Max	Min	Max	Min	Max	Min
4K	13	10.4	30	28	40	35
2K	7.549	5.816	27.5	25.5	30	25
FHD	3.676	2.804	25	23	20	15
HD	2.674	1.229	22.5	20	10	8

resulting delta aggregated data rate ($\Delta ADR = ADR_{t+1} - ADR_t$) must be as low as possible. DASA must guarantee that available resources later of accepting a new user, in an overload situation, be zero or close to zero. The residual value is defined by

$$Ꞓ = \begin{cases} \frac{\sum_{m\in\mathbb{M}_u} Th^u_{b,m,o}}{P_r}, & if\ P_r \geq Ƒ \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

where $P_r = \sum_{z\in\mathbb{U}} Th^z_r$ is the total released resources to guarantee at least the minimum requirement (i.e., Ƒ). Additionally, $\sum_{m\in\mathbb{M}_u} Th^u_{b,m,o}$ is the total assigned Th to the new user. The $Ꞓ = 1$ means that the algorithm frees up precisely what the new user needs. As we have mentioned, DASA never assigns more resources than the maximum required by the services.

- The characteristic function $\upsilon(S_s)$ plays a critical role because it quantifies the cooperation cost in terms of residual value and number of affected users. In this case, $\upsilon(S_s) \in [0,1]$, and it is defined by

$$\upsilon(S_s) = \begin{cases} \frac{3}{4} * Uaff_N + \frac{1}{4} * Ꞓ, & if\ P_r \geq Ƒ \\ 0, & \text{otherwise} \end{cases} \quad (22)$$

where $Uaff_N$ is the normalized value of affected users represented by

$$Uaff_N = 1 - \frac{U_{aff}}{U}. \quad (23)$$

The cheapest winning coalition is the coalition that satisfies the new service petition with the minimal number of affected users and minimum residual. The characteristic function ponderates a little more the importance of affecting the fewest users.

- Rule of order ρ to determine the sequential game of coalition formation. It is developed according to the following fairness assumptions:

1- It considers two priorities: the services' preferences (PS or SSs) and the tariff plan. The former means that the services with low preference have a minor impact (cost). The latter means that users with low priority pay less and, therefore, they are the first to experience a reduction in their resources.

2- First in, first out (FIFO) policy is also applied. The users that stay longer in the network reduce first their resources.

3- The throughput is sequentially reduced to avoid an abruptly Th_{sat} affectation following the set of actions.

According to the above assumptions, all served users by the players are ranked by priority, the number of used services, and time on the network. Therefore, regular users who have been on the network longer using the largest number of services (four in this case) will be chosen for the game initiation.

When all users' services with low priority have the minimum throughput possible, and the overload situation persists, the throughput of users with high priority decreases, always maintaining superior Th_{sat} concerning regular users. They pay more and therefore demand a better perception according to the SLA.

In order to apply the collaboration among the NSs (four players), we define 15 combinations of possible coalitions: from the NSs acting alone ($S_1 = \{VI\}, \dots, S_4 = \{IoT\}$) to the $S_{15} = \{GM, AR, VI, IoT\}$, where S_{15} is the grand coalition (S_{GC}) of the four NSs.

The coalition formation game is generally non-superadditive. It means that we cannot guarantee that always the $\upsilon(S_{15})$ value is higher than or equal to any other coalition in $\mathbb{S}$. Therefore, we consider a dynamic coalition formation iterating over $\mathbb{S}$ to find the coalition that better satisfies the new services request at each time. This mechanism ensures a stable condition freeing up resources with the minimum number of affected users and the minimum residual value.

The game is finite, and it ends when all services have the minimum throughput possible, and it is unfeasible to free up more resources and accept new users. This condition is a transitory state until resources are released (i.e., a current user leaves the network).

The above collaboration among NSs is described in Algorithm 3.

VI. Simulation and Scenarios Deployment

This section presents the simulation method, the scenarios deployment, and the DASA algorithm's validation process. The simulation framework is based on a dual tool:

- OMNeT++ combined with Simu5G and INET [60] libraries are used as a network-level simulator to evaluate the network configuration and obtain the periodical QoS attributes.

Algorithm 3: Collaboration among NSs

Result: Return the cheapest winning coalition and released resources.
Input: ActionSpace = List of users and services;
Ƒ= Minimum required resources;
$\mathbb{S} = \{S_1\{\text{GM}\}, S_2\{\text{AR}\}, S_3\{\text{VI}\}, S_4\{\text{IoT}\}, S_5\{\text{GM, AR}\}, \dots, S_{11}\{\text{GM, AR, VI}\}, \dots, S_{15}\{\text{GM, AR, VI, IoT}\}\}$;

```
begin
  υ_win = 0;
  P_r = 0;
  U_aff = 0;
  for each s in 𝕊 do
    for each z in ActionSpace do
      P_r = P_r + Th_r^z
      U_aff = U_aff + 1
      if P_r ≥ Ƒ then
        υ(S_s) = 3/4 * Uaff_N + 1/4 * Ꞓ
        if υ(S_s) > υ_win then
          υ_win = υ(S_s)
        end
      end
    end
    if P_r < Ƒ then
      υ(S_s) = 0
    end
  end
end
```

- Python is used to implement DASA and evaluate its performance. The resulting data from OMNeT++ simulations are inputs to this algorithm.

The simulated scenario is shown in Fig. 5. It recreates a smart city with a heterogeneous environment composed of one IEEE 802.11ac AP (BS_1), one micro-cell (BS_2) and one macro-cell 5G NR (BS_3). The AP and the micro-cell coverage radius is 250 m, whereas for the macro-cell the radius is 500 m. We simulate both technologies for 40 MHz of bandwidth in the frequencies of 5 GHz and 3.5 GHz for the 802.11ac and 5G NR, respectively. Table IX shows the main simulation parameters.

Table X describes the NSs accessible by the BSs and their respective states ($St_{b,m}$). Initially, we assume only the $St_{b,m}$ equal to 0 or 1.

In the simulations, we recreate U users randomly positioned in the 800x600 meters of recreated heterogeneous network environment. The BS_1 is ubicated in the position (250, 250), the BS_2 in (550, 350), and the BS_3 in (400, 300). The number of users and their mobility behavior vary according to the recreated scenario.

We assume users with different device characteristics and tariff plans (45% of premium clients and 55% of regular clients). They can request from one to four services: VI, AR, IoT, and GM. According to user preferences, one service is the primary service, and the rest are secondary services. We assume these users' preferences as an input of the algorithm. Then, DASA dynamically calculates the best candidate network for each user, and it assigns the resources considering the priority and network conditions.

The main elements of the DASA validation process can be found in Fig. 6. We set up the network-level simulation in OMNeT++ according to the defined scenario and network parameters. Processing the simulations output files, we use the network parameters, the QoS attributes, and the device profile as inputs of the DASA algorithm. Then we recreate the NS Repository information, the user, and the application profiles conforming to the inputs to run the DASA algorithm and evaluate its performance.

The validation consists of evaluating two major processes: the network selection and the load balancing mechanisms through two scenarios presented below.

A. Scenario 1

Scenario 1 aims at evaluating the DASA network selection

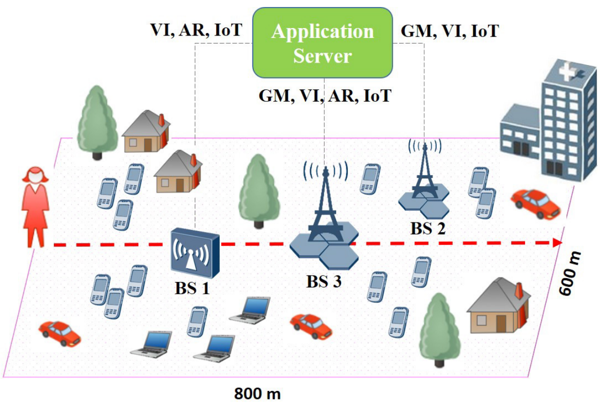

Fig. 5. Simulated heterogeneous network environment.

TABLE IX.
SIMULATION PARAMETERS

Attributes	Cellular Network		WLAN Network
Technology	5G NR		802.11 ac
Number of Tx	2		1
Frequency	3.5 GHz		5 GHz
Transmission Power	macro-cell 25 dBm	micro-cell 10 dBm	10 dBm
Radius Cell	500 m	250 m	250 m
Bandwidth	40 MHz		
Data rate	100-500 Mbps		600-800 Mbps
User type	Pedestrians		
Mobility type	Stationary/RWP/ Linear mobility (1 m/s)		
Channel Model	Rayleigh		

TABLE X.
NETWORK SLICES STATES ($St_{b,m}$)

NS/BS	BS_1	BS_2	BS_3
VI	1	1	1
AR	1	0	1
IoT	1	1	1
GM	0	1	1

process. We consider 50 background users randomly positioned in the area mentioned above around the three BSs, and with a mobility type random waypoint (RWP) [61]. They can request from one to four services simultaneously. Moreover, we recreate a new regular user (UE_{51}) walking through a linear path of 800 m from position (0, 300) to (800, 300) at 1 m/s, as shown in Fig. 5. The user UE_{51} changes the service petition every 100 seconds to analyze the performance of the algorithm under different conditions. The duration of the simulation is 800 seconds to allow the user to complete the route.

This scenario evaluates the dynamic behavior of the network selection process leading to the constant changes of the QoS attributes of the mobile users. The algorithm must ensure to all users the maximum throughput value according to their priorities and network conditions. The traffic generated by the background users significantly influences the available resources for the user walking along the route, especially at the extremes of the BSs' coverage area due to the relatively low signal level.

B. Scenario 2

In this scenario, the number of users has been progressively increased to evaluate the network selection and load balancing mechanisms. The network hosts a new user every two seconds, and the process has a duration of 400 seconds for a total population of 200 users. The users are randomly positioned in the area mentioned above around the three BSs and with a mobility type stationary or RWP.

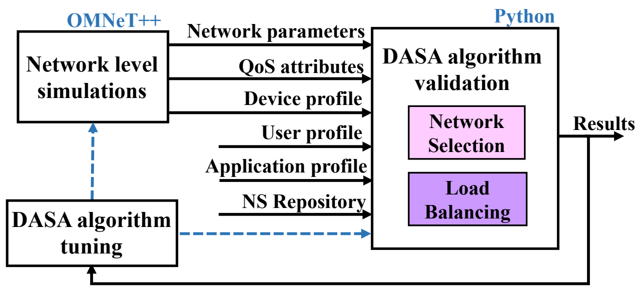

Fig. 6. Elements of the DASA algorithm validation process.

When there are enough resources, DASA maintains all users with the maximum Th_{sat} value according to network conditions and application profile. On the contrary, when the first saturation point occurs, the algorithm performs the load balancing mechanism. The users with low priority experience a gradual reduction in their throughput until they have the minimum possible value. In this second relevant point, DASA decreases the resources from users with high priority accepting more clients and maintaining an adequate QoS according to the SLA. Finally, when all users have the minimum possible resources, the game ends, and it is unfeasible to accept more clients.

VII. Results and Discussion

This section shows the results from the simulated scenarios (Section VI), validating the DASA algorithm performance in a heterogeneous environment for the network selection and management of resource allocation during an overload situation. For simplicity, in the figures, the term game theory appears as GT.

A. *Results of Scenario 1*

Scenario 1 evaluates the Th_{sat} and the dynamic network selection behavior for the user UE_{51} along the route, under the influence of dynamic background traffic. The generated figures show the results of a single simulation run.

DASA is compared with three state of the art methods: the traditional RSS criterion, TYDER [10], and Max-SINR [33]. TYDER calculates the network reputation in terms of QoS parameters to select the most suitable BS that satisfies an individual traffic type requested by the user. Therefore, we assume that TYDER only guarantees the best conditions for the PS.

The Max-SINR method is a variation of the traditional RSS handover process. In this case, we assume that Max-SINR selects the BS with the highest SINR and that supports the major number of requested services, including the PS. The selected BS must guarantee at least the minimum required throughput for the PS.

Fig. 7 shows the access network selection behavior to satisfy the user UE_{51} petition along the route, applying DASA, RSS, TYDER, and Max-SINR. Fig. 8 shows the resulting Th_{sat} values experienced by the user UE_{51} due to the network selection process.

DASA always selects the best combination of access network/NSs, considering the user priority, network conditions, and the NS accessibility at each BS (Table X). Fig. 8 shows how DASA outperforms the other methods providing a maximum Th_{sat} for the user UE_{51}, except at the extremes of the route. At these points of the simulated area, the user petition is GM, VI, and IoT and the BSs cannot guarantee the necessary channel quality and resources blocks to assign the maximum throughput. All the analyzed methods give the same result in this cell-edge area.

On the other hand, the RSS criterion has the worst performance. This method only uses the SINR value for making handover decisions, whereby the target BS might not be able to provide the required services. For example, between

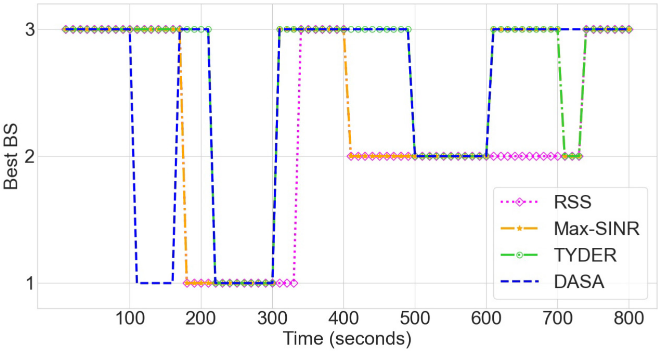

Fig. 7. Best BSs for the user UE_{51} in scenario 1.

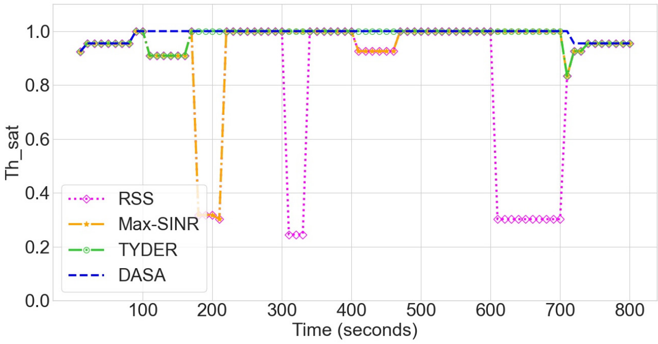

Fig. 8. Th_{sat} values for the user UE_{51} in scenario 1.

300 and 400 seconds of the simulation, the requested services by UE_{51} are GM and VI. In this condition, RSS selects the BS_1 although this BS does not support GM, resulting in the Th_{sat} degradation shown in Fig. 8. Moreover, the Th_{sat} value is considerably affected when RSS selects the BS_2 during the user petition of AR and VI (between 600 to 700 seconds of the simulation), because this BS does not support AR.

Max-SINR shows a better performance than RSS because it considers the accessibility of the requested services. However, albeit the target BS supports the user petition, the achievable QoS metrics may be insufficient. This situation is evident during 180 to 210 seconds of the simulation when RSS and Max-SINR methods select BS_1 without considering QoS performance.

TYDER selects the most suitable BS according to the QoS metrics demanded by the PS. Therefore, during high traffic conditions, this algorithm might not guarantee the maximum Th_{sat} for the rest of the requested services. This is shown, for example, during 110 to 160 seconds of the simulation, when the petition of the user UE_{51} is AR, VI and IoT. TYDER, RSS and Max-SINR select the BS_3, although this BS does not offer the best QoS conditions for all the requested services.

Additionally, Table XI shows the average Th_{sat} for the user UE_{51} along the route resulting from 50 simulation runs with the 95% Confidence Interval (CI). DASA outperforms

TABLE XI.
Average Th_{sat} for the User UE_{51} in Scenario 1

Algorithm	DASA	TYDER	Max-SINR	RSS
Th_{sat} average	0.981	0.943	0.922	0.805
(95% CI)	(0.975-0.987)	(0.928-0.958)	(0.910-0.934)	(0.771-0.839)

TYDER on 3.8%, Max-SINR on 5.9%, and RSS on 17.6%, proving its advantages in terms of dynamic network selection and slice allocation.

B. Results of Scenario 2

Scenario 2 evaluates the network selection and load balancing mechanisms in terms of ADR, Th_{sat}, and SD. The generated figures show the results of a single simulation run. For better understanding, we summarized the evaluated cases in Table XII.

Fig. 9 shows the ADR resulting in applying the collaboration among candidate networks (first phase of GT) and without cooperation. Both cases are analyzed fixing the grand coalition structure ($S_{GC}\{GM, AR, VI, IoT\}$) to reduce the resources and handle the congestion control mechanism. Fig. 10 shows the Th_{sat} standard deviation among the three BSs for the two analyzed cases.

The figures show the improvement of applying the first phase of our game theory proposal. Since the first saturation point is reached, the ADR behavior is more stable around the system's maximum capacity due to the collaboration among BSs. This result means efficient use of the potential released resources. Moreover, the Th_{sat} standard deviation is significantly less when the candidate networks collaborate, proving that one BS is not affected more than others. It is maintained a balance among the resources of the candidate networks following a fairness strategy.

Fig. 11 and 12 show the comparison between three different cases. Two of them apply only the first phase and maintain static the NS coalition: one represents the grand coalition (S_{GC}), and the second is another fixed coalition structure (S_{1-4}). In the case of S_{1-4}, DASA applies the reduction of resources starting from S_1 to S_4, independently. The third case denotes the collaboration among candidate networks and NSs (first and second phases of GT).

The above comparison proves that the grand coalition does not always provide the best performance. Therefore, the superadditive property is not met in the collaboration among NSs. A dynamic coalition selection at each saturation point guarantees the best payoff affecting a smaller number of users and less residual (Ɐ). For 200 users, the two cases of the first phase of GT (fixed S_{1-4} and S_{GC}) affect 187 and 177 users during the load balancing process, respectively. In contrast, the combination of the first and second phases affects 159 users. Additionally, the Ɐ value has the worst performance for the first phase of GT with the fixed S_{1-4} coalition, contrary to the combination of the first and second phases that shows the values closest to 1.

TABLE XII.
EVALUATED CASES IN SCENARIO 2

Cases	1st phase	2nd phase
Static GT	Without GT	Fixed S_{GC}
1st phase GT (fixed S_{GC})	Collaboration among $CANs$	Fixed S_{GC}
1st phase GT (fixed S_{1-4})	Collaboration among $CANs$	Fixed S_{1-4}
1st + 2nd phases GT	Collaboration among $CANs$	Collaboration among NSs

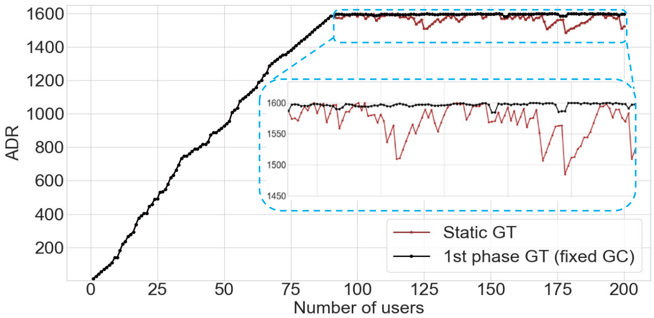

Fig. 9. ADR for the static GT and the first phase of GT with fixed S_{GC}.

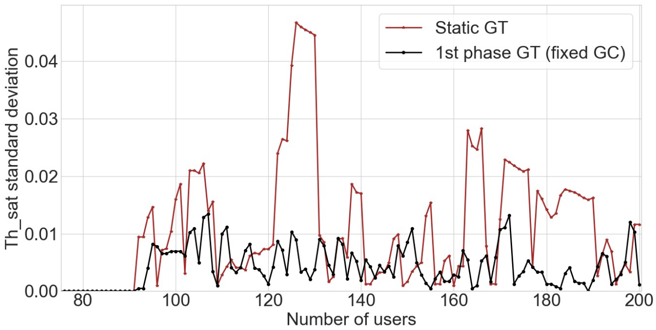

Fig. 10. Th_{sat} standard deviation for the static GT and the first phase of GT with fixed S_{GC}.

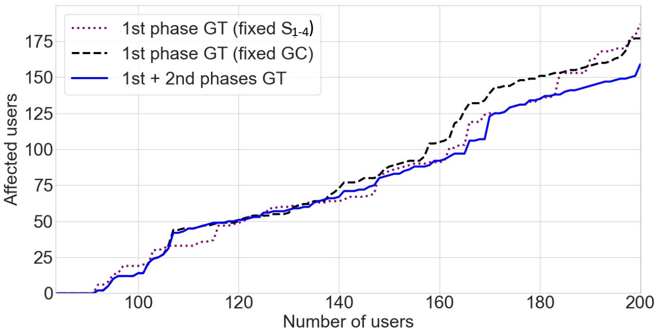

Fig. 11. Affected users for the first phase of GT and the combination of the first and second phases of GT.

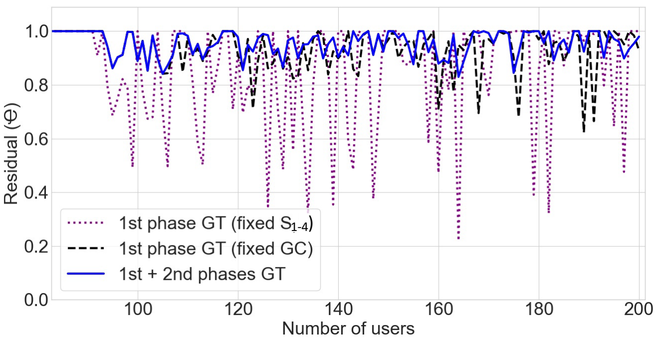

Fig. 12. Residual for the first phase of GT and the combination of the first and second phases of GT.

Once we validate the outperformance of the first and second phases of GT, we extend the population size to reach the full saturation point. Fig. 13 and 14 evaluate the proposal in terms of ADR and average Th_{sat}. It is shown that all users experience the maximum Th_{sat} value until the network reaches the first saturation point. At this point, the regular clients reduce their throughput, whereas the premium clients maintain the best Th_{sat}. When the current regular users do not have enough resources to free up and accept a new user, the

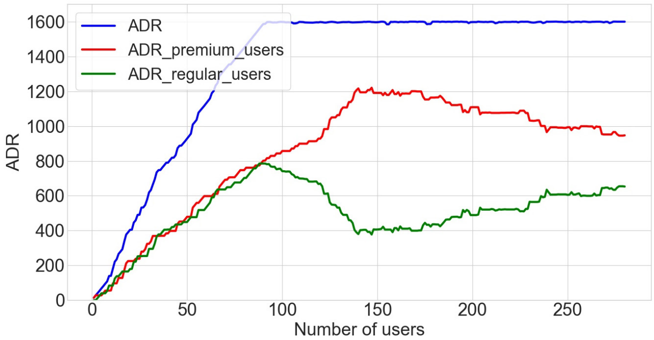

Fig. 13. ADR for the combination of the first and second phases of GT.

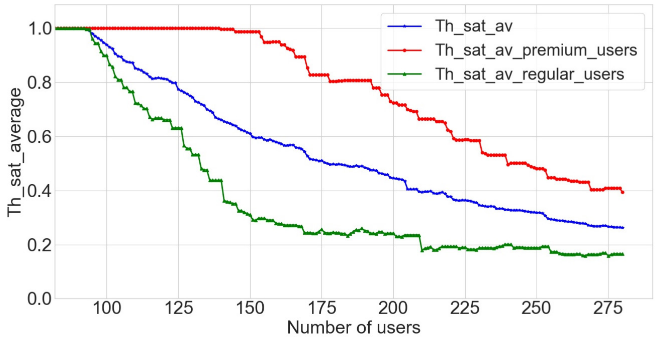

Fig. 14. Average Th_{sat} for the combination of the first and second phases of GT.

premium clients experience a reduction in their resources. If residual resources are available on the network during the congestion control process, affected users can revert to the previous state.

For this scenario and this specific simulation run, the full saturation occurs with 280 users. At this point, all users have the minimum QoS according to their priorities. The game ends, and any new service request is rejected unless new resources are released. Fig. 14 shows that users with high priority always maintain better QoS concerning users who pay less.

Fig. 15 illustrates the average Th_{sat} behavior for each BS. It demonstrates that the candidate network with the highest Th_{sat} is always selected at each saturation point, except when the BS does not support the requested services. Note that when UE_{137} requests GM, the selected candidate network is BS_2 because even though BS_1 has higher Th_{sat}, it does not support the service petition.

Finally, Table XIII summarizes the average values of Th_{sat} and SD for a total of 200 users resulting from 50 simulation

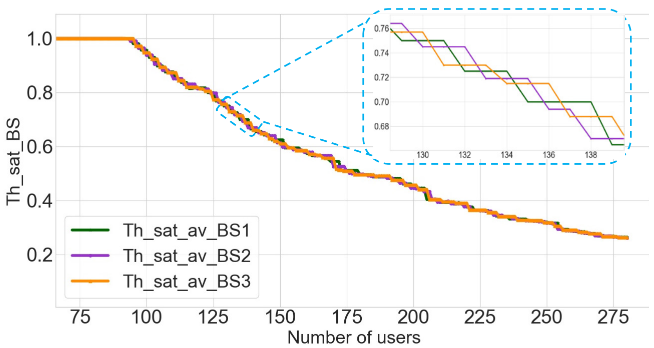

Fig. 15. Average Th_{sat} of the BSs for the combination of the first and second phases of GT.

TABLE XIII.
RESULTS OF SCENARIO 2 FOR 200 USERS

	Static GT	1st phase (fixed S_{1-4})	1st phase (fixed S_{GC})	1st + 2nd phase
Average Th_{sat}	0.394	0.411	0.420	0.459
(95% CI)	(0.387-0.401)	(0.404-0.418)	(0.406-0.434)	(0.452-0.466)
Average Th_{sat} (premium users)	0.687	0.680	0.697	0.713
(95% CI)	(0.675-0.699)	(0.647-0.713)	(0.673-0.722)	(0.691-0.735)
Average Th_{sat} (regular users)	0.158	0.210	0.211	0.243
(95% CI)	(0.148-0.169)	(0.177-0.243)	(0.172-0.250)	(0.223-0.263)
Average SD	0.915	0.957	0.969	0.987
(95% CI)	(0.905-0.925)	(0.948-0.966)	(0.959-0.979)	(0.983-0.991)
Average SD (premium users)	0.960	0.962	0.973	0.988
(95% CI)	(0.946-0.974)	(0.956-0.968)	(0.960-0.986)	(0.983-0.993)
Average SD (regular users)	0.865	0.952	0.966	0.987
(95% CI)	(0.845-0.885)	(0.940-0.964)	(0.959-0.973)	(0.982-0.992)

runs with the 95% CI. The combination of the first and second phases improves the average Th_{sat} regarding the static GT case and the first phase of GT with the fixed coalitions structures S_{1-4} and S_{GC} by 6.5%, 4.8%, and 3.9%, respectively. At the same time, the combination of the first and second phases of GT improves the average SD by 7.2%, 3%, and 1.8%, respectively. These results prove the advantages of applying two collaborative phases of GT among BSs and among NSs.

VIII. CONCLUSION

DASA is an integral solution to handle the dynamic slice allocation during the network selection and overload situation. It combines the NS paradigm with SDN and NFV as crucial technologies to introduce flexibility in traffic management for future wireless networks.

The proposed algorithm is based on MADM and AHP to combine multiple attributes and face the complex problem of network selection over 5G heterogeneous networks and beyond. DASA treats and differentiates the critical services VI, AR, IoT, and GM in terms of throughput, delay, jitter, packet loss ratio, and energy consumption. Moreover, the proposal considers users with different tariff plans, device resolutions, and service preferences. The above assumptions guarantee to select the most efficient combination of access network/NSs to satisfy the user petition.

On the other hand, DASA applies a cooperative game theory approach to drive the load balancing during an overload situation. It considers two cooperative phases among BSs and among NSs, ensuring to maintain a balance in the Th_{sat} and to accept more users in the network. The premium clients always are benefited over the regular ones, which is evident in the simulation results.

The algorithm is evaluated through network-level simulations integrating OMNeT++ with Simu5G and Python tools. It is demonstrated the outperformance of the proposed dynamic network selection algorithm through Th_{sat} values concerning RSS, TYDER, and Max-SINR methods. Finally, combining two phases of collaboration during the load balancing mechanism improves the algorithm performance, avoiding the abrupt throughput reduction and affecting a small number of users at each saturation point.

REFERENCES

[1] U. Cisco, "Cisco annual internet report (2018–2023) white paper," ed, 2020.

[2] ETSI, "Technical report: Speech and multimedia Transmission Quality (STQ); QoS parameters and test scenarios for assessing network capabilities in 5G performance measurements " in *TR 103 702 V1.1.1*, ed, 2020, p. 45.

[3] G. Mayer, "Standards for the Global Internet of Things. How 3GPP Enables Massive Machine Type Communications," presented at the ETSI IoT Workshop 2019, 2019.

[4] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine,* vol. 58, pp. 55-61, 2020.

[5] T. Huang, W. Yang, J. Wu, J. Ma, X. Zhang, and D. Zhang, "A survey on green 6G network: Architecture and technologies," *IEEE Access,* vol. 7, pp. 175758-175768, 2019.

[6] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society,* vol. 1, pp. 957-975, 2020.

[7] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Communications Magazine,* vol. 55, pp. 94-100, 2017.

[8] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, "5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges," *Computer Networks,* vol. 167, p. 106984, 2020.

[9] F. H. Fitzek, F. Granelli, and P. Seeling, *Computing in Communication Networks: From Theory to Practice*: Academic Press, 2020.

[10] C. Desogus, M. Anedda, M. Murroni, and G.-M. Muntean, "A traffic type-based differentiated reputation algorithm for radio resource allocation during multi-service content delivery in 5G heterogeneous scenarios," *IEEE Access,* vol. 7, pp. 27720-27735, 2019.

[11] J. Montalban, G.-M. Muntean, and P. Angueira, "A utility-based framework for performance and energy-aware convergence in 5G heterogeneous network environments," *IEEE Transactions on Broadcasting,* vol. 66, pp. 589-599, 2020.

[12] A. Alinezhad, Khalili, Javad, *New Methods and Applications in Multiple Attribute Decision Making (MADM)*, 2019.

[13] M. Anedda, M. Murroni, and G.-M. Muntean, "A novel Markov decision process-based solution for improved quality prioritized video delivery," *IEEE Transactions on Network and Service Management,* vol. 17, pp. 592-606, 2019.

[14] C. Desogus, M. Anedda, M. Fadda, and M. Murroni, "Additive Logarithmic Weighting for Balancing Video Delivery Over Heterogeneous Networks," *IEEE Transactions on Broadcasting,* 2020.

[15] R. J. Boucherie and N. M. Van Dijk, *Markov decision processes in practice*: Springer, 2017.

[16] Z. Han, D. Niyato, W. Saad, and T. Başar, *Game theory for next generation wireless and communication networks: Modeling, analysis, and design*: Cambridge University Press, 2019.

[17] (Feb. 16). *OMNeT++ Discrete Event Simulator*. Available: https://omnetpp.org/

[18] G. Nardini, D. Sabella, G. Stea, P. Thakkar, and A. Virdis, "Simu5G–An OMNeT++ Library for End-to-End Performance Evaluation of 5G Networks," *IEEE Access,* vol. 8, pp. 181176-181191, 2020.

[19] H. Zhang and W. Quan, "Networking automation and intelligence: A new era of network innovation," *Engineering,* 2021.

[20] T. Taleb, I. Afolabi, K. Samdanis, and F. Z. Yousaf, "On multi-domain network slicing orchestration architecture and federated resource control," *IEEE Network,* vol. 33, pp. 242-252, 2019.

[21] Z. Shu and T. Taleb, "A novel QoS framework for network slicing in 5G and beyond networks based on SDN and NFV," *IEEE Network,* vol. 34, pp. 256-263, 2020.

[22] F. Z. Yousaf, M. Gramaglia, V. Friderikos, B. Gajic, D. von Hugo, B. Sayadi, *et al.*, "Network slicing with flexible mobility and QoS/QoE support for 5G Networks," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2017, pp. 1195-1201.

[23] T. Ahmed, A. Alleg, R. Ferrus, and R. Riggio, "On-demand network slicing using SDN/NFV-enabled satellite ground segment systems," in *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*, 2018, pp. 242-246.

[24] M. Chahbar, G. Diaz, A. Dandoush, C. Cérin, and K. Ghoumid, "A Comprehensive Survey on the E2E 5G Network Slicing Model," *IEEE Transactions on Network and Service Management,* 2020.

[25] I. Mashal, O. Alsaryrah, T.-Y. Chung, and F.-C. Yuan, "A multi-criteria analysis for an internet of things application recommendation system," *Technology in Society,* vol. 60, p. 101216, 2020.

[26] R. V. Rosa and C. E. Rothenberg, "The pandora of network slicing: A multicriteria analysis," *Transactions on Emerging Telecommunications Technologies,* vol. 31, p. e3651, 2020.

[27] B. Bakmaz, Z. Bojkovic, and M. Bakmaz, "TOPSIS‐based approach for network slice selection in 5G mobile systems," *International Journal of Communication Systems,* vol. 33, p. e4395, 2020.

[28] D. L. Olson, "Comparison of weights in TOPSIS models," *Mathematical and Computer Modelling,* vol. 40, pp. 721-727, 2004.

[29] G. Zhao, S. Qin, G. Feng, and Y. Sun, "Network slice selection in softwarization‐based mobile networks," *Transactions on Emerging Telecommunications Technologies,* vol. 31, p. e3617, 2020.

[30] D. Whitley, "A genetic algorithm tutorial," *Statistics and computing,* vol. 4, pp. 65-85, 1994.

[31] A. Ahmed, L. M. Boulahia, and D. Gaiti, "Enabling vertical handover decisions in heterogeneous wireless networks: A state-of-the-art and a classification," *IEEE Communications Surveys & Tutorials,* vol. 16, pp. 776-811, 2013.

[32] A. Vince, "A framework for the greedy algorithm," *Discrete Applied Mathematics,* vol. 121, pp. 247-260, 2002.

[33] Y. Sun, W. Jiang, G. Feng, P. V. Klaine, L. Zhang, M. A. Imran*, et al.*, "Efficient handover mechanism for radio access network slicing by exploiting distributed learning," *IEEE Transactions on Network and Service Management,* vol. 17, pp. 2620-2633, 2020.

[34] R. Su, D. Zhang, R. Venkatesan, Z. Gong, C. Li, F. Ding*, et al.*, "Resource allocation for network slicing in 5G telecommunication networks: A survey of principles and models," *IEEE Network,* vol. 33, pp. 172-179, 2019.

[35] B. Han, A. DeDomenico, G. Dandachi, A. Drosou, D. Tzovaras, R. Querio*, et al.*, "Admission and congestion control for 5g network slicing," in *2018 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2018, pp. 1-6.

[36] S. K. Singh, M. M. Salim, J. Cha, Y. Pan, and J. H. Park, "Machine learning-based network sub-slicing framework in a sustainable 5g environment," *Sustainability,* vol. 12, p. 6250, 2020.

[37] D. Ferreira, A. Reis, C. Senna, and S. Sargento, "A Forecasting Approach to Improve Control and Management for 5G Networks," *IEEE Transactions on Network and Service Management,* 2021.

[38] A. Baumgartner, T. Bauschert, A. M. Koster, and V. S. Reddy, "Optimisation models for robust and survivable network slice design: A comparative analysis," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, 2017, pp. 1-7.

[39] G. Wang, G. Feng, T. Q. Quek, S. Qin, R. Wen, and W. Tan, "Reconfiguration in network slicing—Optimizing the profit and performance," *IEEE Transactions on Network and Service Management,* vol. 16, pp. 591-605, 2019.

[40] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE network,* vol. 34, pp. 134-142, 2019.

[41] J. Antoniou, *Game Theory, the Internet of Things and 5G Networks*: Springer, 2020.

[42] Z. Han, D. Niyato, W. Saad, T. Başar, and A. Hjørungnes, *Game theory in wireless and communication networks: theory, models, and applications*: Cambridge university press, 2012.

[43] W. A. Gamson, "A theory of coalition formation," *American sociological review,* pp. 373-382, 1961.

[44] A. Lieto, I. Malanchini, S. Mandelli, E. Moro, and A. Capone, "Strategic network slicing management in radio access networks," *IEEE Transactions on Mobile Computing,* 2020.

[45] *Recommendation Y.3150 (09/20): High-level technical characteristics of network softwarization for IMT-2020* 2020.

[46] *Y.3111 (09/2017): IMT-2020 network management and orchestration framework* 2017.

[47] G. D. P. Regulation, "General data protection regulation (GDPR)," *Intersoft Consulting, Accessed in October,* vol. 24, 2018.

[48] ETSI, "TS 138 300 V15.8.0 5G; NR; Overall description; Stage-2," ed, 2020, p. 102.
[49] Q.-T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, "On utility models for access network selection in wireless heterogeneous networks," in *NOMS 2008-2008 IEEE Network Operations and Management Symposium*, 2008, pp. 144-151.
[50] S. Radouche, C. Leghris, and A. Adib, "MADM methods based on utility function and reputation for access network selection in a multi-access mobile network environment," in *2017 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2017, pp. 1-6.
[51] R. Khoder, R. Naja, N. Mouawad, and S. Tojme, "Vertical Handover Network Selection Architecture for VLC Vehicular Platoon Driving Assistance," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1-6.
[52] G. Sutorius, "Jitter Separation in High Speed Digital Design," Agilent Technologies2018.
[53] J. Ding, M. Nemati, C. Ranaweera, and J. Choi, "IoT connectivity technologies and applications: A survey," *arXiv preprint arXiv:2002.12646,* 2020.
[54] R. W. Saaty, "The analytic hierarchy process—what it is and how it is used," *Mathematical modelling,* vol. 9, pp. 161-176, 1987.
[55] A. B. MacKenzie and L. A. DaSilva, "Game theory for wireless engineers," *Synthesis Lectures on Communications,* vol. 1, pp. 1-86, 2006.
[56] Huawei, "Empowering consumers focused immersive VR and AR experience with mobile broadband," 2016.
[57] A. Wang, L. Yu, A. Mudesir, D. Zhu, B. Zhao, and T. Siew, "5G Unlocks a World of Opportunities: Top Ten 5G Use Cases," *Huawei Technologies Co., Shenzen, China,* 2017.
[58] Qualcomm, "VR and AR pushing connectivity limits," 2018.
[59] V. VYAS. (2020). *Cloud Gaming Explained: Could It Be the Next Big Change?* Available: https://arcanelost.com/cloud-gaming-explained-next-big-change/
[60] (2021). *The INET Framework*. Available: https://inet.omnetpp.org/
[61] E. Hyytiä and J. Virtamo, "Random waypoint mobility model in cellular networks," *Wireless Networks,* vol. 13, pp. 177-188, 2007.

Claudia Carballo González received the B.Sc. degree (summa cum laude) in Telecommunications and Electronics Engineering, and the M.Sc. degree in Telecommunications and Telematics from the Havana University of Technologies (CUJAE) in 2015 and 2020, respectively. She is currently a Ph.D. student at the University of Cagliari, with the Department of Electrical and Electronic Engineering (DIEE/UdR CNIT), under the supervision of Prof. M. Murroni. Her research interests include QoS/QoE, 5G wireless networks and beyond, resource allocation, and artificial intelligence.

Ernesto Fontes Pupo received the B.Sc. degree in Telecommunications and Electronics Engineering and the M.Sc. degree in Digital Systems from the Havana University of Technologies (CUJAE) in 2014 and 2018, respectively. He is currently a Ph.D. student at the University of Cagliari, with the Department of Electrical and Electronic Engineering (DIEE/UdR CNIT), under the supervision of Prof. M. Murroni. He is also an assistant researcher at LACETEL, R&D Telecommunications Institute (2014-current). His research interests include multimedia broadcast/multicast, 5G wireless networks and beyond, physical layer development, and artificial intelligence.

Luigi Atzori (Ph.D., 2000) is a professor of Telecommunications at the University of Cagliari, where he leads the activities of the MCLab laboratory (Multimedia & Communications) with around 20 affiliated researchers. Since 2018, he has been the coordinator of the master degree course in Internet Technology Engineering at the University of Cagliari. His research interests fall in the area of Internet of Things (IoT), with particular reference to the design of effective algorithms for the realization of social networks among connected devices to create the Social IoT paradigm. His interests also fall in the area of Quality of Experience (QoE), with particular application to the management of services and resources in new generation networks for multimedia communications. Lately, he also applies the study of QoE to IoT services. He serves regularly in the conference organizing committee of the sector and as an associate and guest editor of several international journals (IEEE IoT journal, Ad Hoc Networks, IEEE Open Journal of the Communications Society, IEEE Communications Magazine, etc.).

Maurizio Murroni, (M. '02, SM. '13) is an associate professor of Telecommunications at the Department of Electrical and Electronic Engineering (DIEE) of the University of Cagliari. He is a member of the National Interuniversity Consortium for Telecommunications (CNIT). He received an M.Sc. in Electronic Engineering in 1998 and a Ph.D. in Electronic and Computer Engineering in 2001 at the University of Cagliari. He was visiting scholar at the School of Electronic Engineering, Information Technology and Mathematics, University of Surrey, Guildford, UK in 1998 and at the Image Processing Group at Polytechnic University, Brooklyn, New York, USA, in 2000. In 2006 and 2014 he was visiting professor at the Department of Electronics and Computers of the University of Transylvania in Brasov in Romania, in 2011 at the Department of Electronics and Telecommunications, Faculty of Engineering of Bilbao, University of the Basque Country (UPV / EHU) in Spain and in 2019 at Kingston University in London, UK. In 2016 he founded the Italian chapter of the IEEE Broadcast Technology Society (BTS). Since 2018 he is a distinguished lecturer for BTS and associate editor for IEEE Transactions on Broadcasting. Prof. Murroni is co-author of an extensive list of articles published in journals and on international conference proceedings and has received numerous awards and recognitions. He has served as chair for various international conferences and workshops. He is co-author of 1900.6-2011 - IEEE Standard for Spectrum Sensing Interfaces and Data Structures for Dynamic Spectrum Access and other Advanced Radio Communication Systems. His current research focuses on broadcast/multicast delivery on 5G networks and beyond, QoE, HbbTV, multi sensorial media.