Control and Management of Multiple RATs in Wireless Networks: An SDN Approach

Akshatha Nayak Manjeshwar[1], Arghyadip Roy[1], Pranav Jha[1], and Abhay Karandikar [1,2]
[1]Department of Electrical Engineering, Indian Institute Technology Bombay
Email: {akshatha, arghyadip, pranavjha, karandi}@ee.iitb.ac.in
[2]Director and Professor, Department of Electrical Engineering, Indian Institute Technology Kanpur
Email: {karandi}@iitk.ac.in

***Abstract*—Multiple Radio Access Technologies (RATs) co-exist within today's mobile broadband networks, and each of these RATs is controlled by a different set of entities, leading to fragmented network control. This may lead to sub-optimal utilization of the overall network resources. In this paper, we propose a novel Software Defined Networking (SDN) based network architecture for unified control of multiple RATs and provide a framework for improved network performance over those of the present-day architectures. The proposed architecture enables end-to-end network control, while preserving scalability with the help of network slicing. We develop an evaluation platform based on network simulator-3 (ns-3) in accordance with the proposed architecture. We also demonstrate the performance improvements provided by the proposed architecture through simulations.**

I. Introduction

A multitude of Radio Access Technologies (RATs) exist in today's wireless networks. However, each of these RATs is controlled by one or more RAT-specific entities. For example, Access Points (APs) in Wireless Local Area Networks (WLANs) are controlled by Access Controllers (ACs) [1], whereas the Fourth Generation (4G) Long Term Evolution (LTE) RAT is controlled by entities such as Mobility Management Entity (MME) and eNodeB (eNB). Fragmentation of the control plane in a multi-RAT network prohibits a global view of the network resources. This lack of a global view hinders the optimized allocation of resources [2]. Even in the upcoming Third Generation Partnership Project (3GPP) Fifth Generation (5G) network which has a common core supporting multiple RATs, radio access related decisions are taken separately within individual RATs [3]. As a result at present, unified control of multi-RAT networks is not supported by 3GPP 5G. By devising mechanisms for unified control and management of multi-RAT networks, we can control and manage diverse RATs in a unified manner. This can be achieved by using the Software Defined Networking (SDN) [4] paradigm.

SDN is a networking principle that decouples the control and data planes [4]. The control plane of a network comprises control and management elements and protocols, whereas the data plane comprises elements/functions that forward data. SDN provides a logical centralization of network control. Furthermore, SDN enables the creation of multiple logical networks known as "network slices" over a common physical infrastructure. Network slicing facilitates the provisioning of one or more logical networks with diverse service requirements over the same physical network.

In this paper, we present an SDN based wireless network architecture which unifies the control and management of diverse RATs. The architecture provides mechanisms for end-to-end control of the wireless network. Although desirable for ease of control and management, centralized control may give rise to scalability issues in large networks. Therefore, the proposed architecture also consists of a network slice orchestrator which splits the end-to-end physical network into multiple logical networks (or network slices) based on the service requirements. Network slicing, with a controller for each network slice, also brings scalability to the architecture. Each slice comprises data plane nodes with an associated control plane entity known as the multi-RAT controller which manages the data plane nodes in a unified manner. The proposed architecture provides a framework for the deployment of RAT-agnostic control applications. It also provides the flexibility to support other future RATs in the integrated framework. In order to evaluate the proposed architecture, we develop a network simulator-3 (ns-3) based evaluation platform in accordance with the proposed architecture. We also illustrate the performance improvements provided by the proposed architecture through experimental results considering slice specific user-association as an example use-case.

A. Existing Architectures for Multi-RAT Network Control

Various attempts have been made for integration of multiple RATs through both research and standardization activities. Works such as [5]–[7] propose two-tiered cloud architectures for the control of multi-RAT networks. Control and management tasks related to mobility, resource allocation and interference are handled by the core cloud, whereas the edge cloud takes care of the RAN functions. The authors in [7] utilize spare bits of the OpenFlow header to implement virtual networks and enable multi-RAT control. This architecture makes use of a higher level network controller for provisioning network nodes and local controllers at the remote radio heads. The authors in [8] present a three-tiered architecture consisting of the physical, control and management layers for dense multi-RAT networks. In [9], the authors present a three-tiered architecture with a flat user plane. This is achieved by encapsulating the protocol layers of the controlled RATs as a module. Another work [10], describes an architecture comprising three clouds, based on the functionalities of the network elements viz., control, access and forwarding. This architecture is realized using Network Function Virtualization (NFV), and service improvement is achieved by placing the user plane functions e.g., Gateways (GWs) closer to the network edge. An approach for unified control and management of multi-RAT networks described in [11], proposes a clean slate architecture

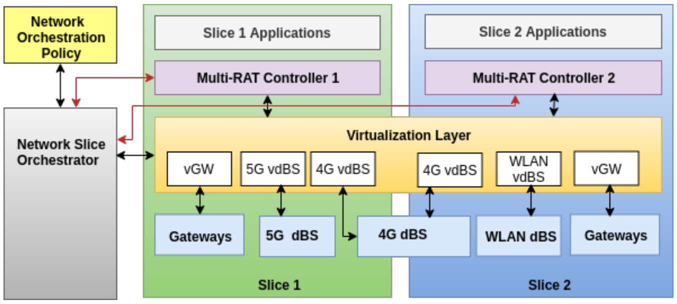


Fig. 1: Block diagram of the proposed network architecture.

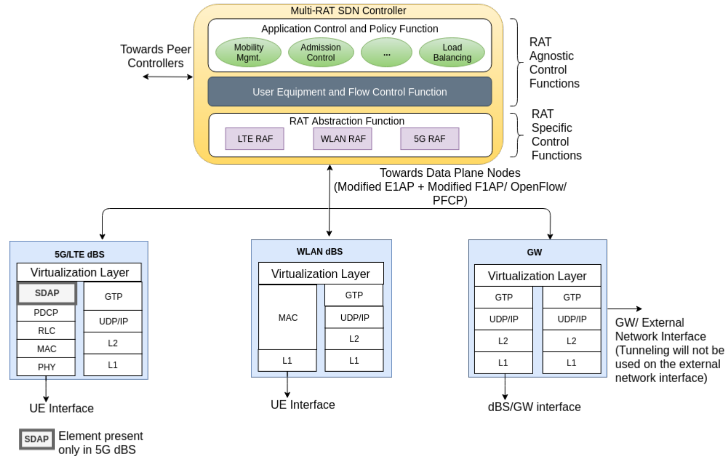


Fig. 2: Architecture of the proposed SDN based multi-RAT network.

known as Cloud-RAN (C-RAN). Within C-RAN, most of the network processing is carried out in the cloud. Our earlier work also explores unified control [12] for wireless networks. Unlike this paper which proposes mechanisms for control of multi-RAT networks, it considers only the 5G NR RAT. Moreover, network slicing aspects are not investigated in [12].

Existing works on slicing in the multi-RAT scenario include [13] and [14]. In [13], slicing is achieved by aggregating the network entities that are shared by different services into common sub-slices which are controlled by a coordinator. The authors of [14] propose a framework to specify and support the creation of RAN slices using configuration descriptors at every layer of the radio protocol stack, i.e., L1, L2, L3 layers of the 5G stack. These descriptors are used to characterize the policies, features, and resources within the protocol layers.

Although a few works in the existing literature are focused on solutions for multi-RAT control, to the best of our knowledge, no other work presents a unified framework for end-to-end multi-RAT network control while ensuring scalability. Our work also abstracts out the RAT-specific details to enable a uniform method for control and management of multi-RAT wireless networks. Other contributions offered by the proposed architecture are as follows:

- A diverse set of requirements for different slices within the network can be handled through the implementation of slice-specific controllers and slice-specific data plane functions.
- Signaling towards the User Equipment (UE) remains broadly unchanged, making this architecture ideal for deployment.
- It also enables features requiring interaction across multiple nodes in a simple manner such as LTE WLAN Aggregation (LWA), multi-connectivity in comparison to existing 3GPP 5G/4G networks.

The rest of the paper is organized as follows. Details of the proposed SDN based architecture are provided in Section II. The succeeding section highlights the advantages of the proposed network architecture. The proposed architecture is then experimentally evaluated in Section IV, followed by conclusions in Section V.

II. Proposed Multi-RAT Network Architecture

The proposed architecture (illustrated in Fig. 1), comprises control entities such as multi-RAT controllers and the network slice orchestrator. The network slice orchestrator creates one or more slices on top of the network infrastructure based on the network orchestration policy. This results in the creation of end-to-end logical networks or network slices by grouping a set of network resources based on service requirements while isolating them from each other. Each slice may consist of a subset of resources from different data plane entities such as data plane Base Stations (dBSs), Gateways (GWs) and is controlled by a control plane entity viz., the multi-RAT SDN controller. The size of a given network slice can be increased or shrunk by re-grouping the physical resources.

The dBSs are RAT-specific data plane entities which are created by eliminating the control functionalities such as radio resource control and management, mobility management, from the respective RAT-specific base stations. For example, as illustrated in Fig. 2, an LTE dBS and a 5G dBS consist of only the forwarding plane of the base station viz., Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC), Medium Access Control (MAC) and Physical Layer (PHY) and an optional virtualization layer. The UE specific dedicated (radio) resource control functionality is moved out of the base stations and placed in the controller. dBSs are responsible for carrying UE specific control and data. The control messages are exchanged with the controller, and data is forwarded to the core network. The cell Radio Resource Control (RRC) functionality which is common to UEs within the cell e.g., provision of configuration for cell broadcast, is pushed to the virtualization layer. Similarly, WLAN dBSs may consist of PHY and MAC layers along with the virtualization layer.

The virtualization layer provides an abstract resource view to the controller which can be based on a virtualization policy. The virtualization layer on an LTE dBS can manifest multiple virtual LTE dBSs (LTE vdBSs) on top of a single physical dBS by partitioning the physical resource blocks available at the LTE dBS and allocating them to the individual virtual dBSs. For example, as illustrated in Fig. 3, 5G dBS1 has been split into two virtual dBSs with 90% of its resources allocated to vdBS1 and the remaining 10% of the resources to vdBS2. Each of these vdBSs are allocated to different slices.

GWs are generic data plane nodes, responsible for forwarding user plane data towards other GWs/external data networks on the uplink and other GWs/dBSs on the downlink. A GW supports data forwarding for all types of UEs and all types of RATs. The virtualization layer may also be present at the GWs, where it manifests virtual gateways (vGWs). The virtualization layer can also be deployed within the network as a separate entity between the controller and the data plane nodes.

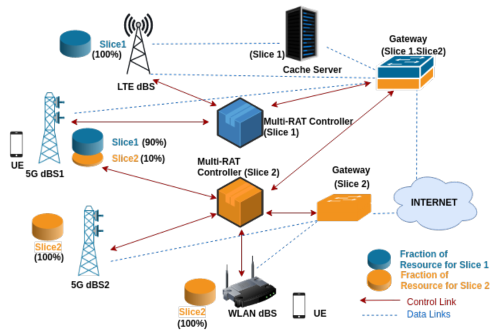


Fig. 3: An example deployment within the proposed architecture.

Each of these virtual data plane entities. e.g., vdBSs or vGWs is a part of a network slice or a logical network. dBSs are responsible for forwarding user plane data exchanged between UEs and external data networks either directly (e.g., when connected to a local cache server) or via the GWs. They are also responsible for forwarding signaling/control plane messages exchanged between the UEs and the controller.

The controller controls and manages the virtual data plane entities within the slice and provides data flow configurations to them. It is also responsible for exchanging control plane messages with the UEs. It may also exchange control plane messages with controllers which are a part of other slices. The communication between slice controllers may be required when a single UE is communicating over multiple slices. In an example scenario (Fig. 3), each network slice has a separate controller which controls the vdBSs and vGWs belonging to that slice. However, controllers may also be shared across slices. Each slice may be governed by specific policies for resource management. Note that the proposed architecture is scalable due to the presence of multiple slices and controllers.

A. *Proposed Multi-RAT Controller Architecture*

Fig. 2 illustrates the architecture of the multi-RAT SDN controller and the data plane nodes within the network. In order to control multiple RATs in a unified manner, functions such as UE authentication, UE mobility management and flow control can be handled in a RAT-agnostic manner. As a result, the controller comprises functionalities for providing RAT-agnostic control and RAT-specific control. The controller comprises the following functions :

• RAT Abstraction Function (RAF): This function is responsible for handling the RAT-specific functionalities within the network. There may exist a separate RAF for every supported RAT. It also manages RAT-specific control plane communication with the UEs. RAF possesses both management and control functionalities and is used to translate generic configuration provided by higher layer functions into RAT-specific configuration to be supplied to a dBS. For example, the 3GPP LTE RAF translates generic flow configuration parameters provided by the layer above into radio bearer parameters to be supplied to an LTE dBS. It also supports the RAT specific Non Access Stratum (NAS) and RRC layers. These layers are responsible for signalling message exchanges with the UE. The rest of the controller modules are RAT-agnostic.

• UE and Flow Control Function (UFCF): UFCF maintains the context for every UE and associates flow(s) with a UE. It is responsible for setting up/handover of flows on dBSs and GWs with the desired Quality of Service (QoS) requirements. UFCF also provides a RAT-independent interface to the layer above which may contain RAT-agnostic control algorithms. UFCF maintains a unified list of abstract attributes such as QoS parameters, UE ID for each connected UE and its associated data flows.

• Application Control and Policy Function (ACPF): ACPF comprises slice-specific control and policy applications. Operators can introduce new applications/policies e.g., admission control, load balancing etc., into a specific slice without affecting other network slices. A RAT-independent interface between the ACPF and the UFCF enables third-party vendors to implement new algorithms.

The southbound interface at the controller can be specified using various protocols that are used to configure the data plane nodes. For example, a modified version of the OpenFlow protocol is used to configure the GW. The 3GPP Packet Forwarding Control Plane protocol (PFCP) protocol [15] may also be used in place of OpenFlow. Similarly, modified versions of E1 and F1 application protocols are used to configure dBSs of 5G, LTE and WLAN RATs.

III. Advantages of the Proposed Architecture

The proposed network architecture offers the following advantages in comparison to the present-day network architecture.

• Unified Authentication and Security: The authentication and security procedures are handled by the controller. Authentication, which is carried out in a unified manner, prevents the need for authenticating the UE every time it connects to a different RAT. This also enables seamless handovers.

• Simplified Signaling procedures: Procedures that require coordination between multiple entities both, within a RAT (e.g., intra-RAT handover) and across multiple RATs (e.g., inter-RAT handover, multi-connectivity) become simpler due to the unified framework for decision making. This is illustrated in Fig. 4 using call flows with 5G NR, 4G LTE and WLAN as the reference RATs. The signaling for 5G NR and 4G LTE is similar to a large extent. We use a common representation for 4G and 5G dBSs as "4G/5G dBS" for the ease of illustration. The decision to perform handover for a UE is taken by the mobility management function of the ACPF within the controller. The measurement reports from the UE are forwarded to the controller to assist in the handover decision. A handover command is sent by the multi-RAT controller to initiate the handover. After the handover, UE is associated with the 4G/5G dBS. Since the UE context is maintained at the controller, re-authentication may not be required. Also, the decision making at multiple individual nodes such as the source and target dBSs, as done in the existing wireless networks, is no longer needed. As a result, multiple handshaking signalling exchanges between the dBSs and the core network used for choosing the target dBS can be eliminated in the proposed architecture. Note that as illustrated in Fig. 4, the signalling exchanges towards the UE remain unchanged.

• Reduced risk of handover failures: Although it has not been

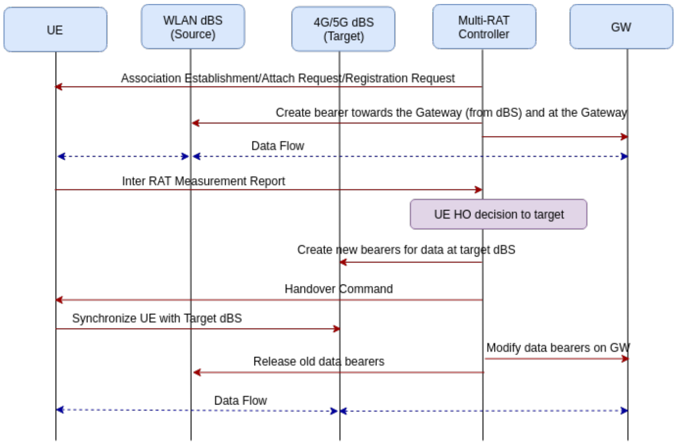


Fig. 4: UE handover (WLAN to 3GPP 4G/5G) call flow within the proposed architecture.

illustrated in the call flow in Fig. 4, a dual connection from the source and target dBSs can be maintained towards the UE during handover. This prevents "ping-pong" handovers as well as reduces the risk of handover failures. It also helps in ensuring session continuity.

• Energy efficiency and power control: Unlike in present-day multi-RAT networks, the SDN controller can regulate power levels for the entire system, thus reducing the overall interference in the RAN. This unified interference management may result in better system throughput. Some dBSs can even be turned off during periods of low traffic by re-distributing the load to the active base stations for increased energy saving.

• Content caching and delivery: By inspecting packets at the controller, data request for popular content can be retrieved from locations near the dBSs instead of the external network through the GW. This results in reduced content retrieval time as well as efficient backhaul usage. Additionally, the source dBS may itself act like an anchor point and continue to serve the UE even after its handover to another dBS.

IV. Experimental Evaluation of the Proposed Architecture

We evaluate the performance offered by the proposed architecture by comparing it with that of the existing network. For this purpose, we develop an evaluation platform[1] based on ns-3 in accordance with the proposed network architecture. The evaluation platform consists of a node with the functionality of the multi-RAT controller. All the network control and management applications are deployed over this node. LTE and WLAN base stations in ns-3 are modified to function as data plane nodes. This is done by removing the UE RRC from the LTE eNB and routing all the control plane messages from the eNB to the controller. On the WLAN AP, the control functionality is removed from the AP by forwarding the messages required for admission control such as association request, from the MAC layer of the AP to the controller. The data plane packets are routed as per the routes configured by the controller. The platform also enables the creation of network slices which are isolated from each other as required by 3GPP specifications [3].

A. Simulation Setup

We create a network slice for best effort traffic in the simulator. The slice consists of a multi-RAT controller that manages an LTE dBS, a WLAN dBS inside the coverage area of the LTE dBS and a GW. We also assume that the GW has enough capacity to support all the best-effort users and hence, does not create a bottleneck. We assume that the users can be associated with either the LTE dBS or the WLAN dBS. However, users present outside the coverage of WLAN dBS are always associated with the LTE dBS. We also assume that the slice consists of a certain fraction of the total capacity of the LTE dBS as specified by the network orchestration policy and a WLAN dBS. Let, C_L be the maximum capacity (in Mbps) provided on the LTE dBS for the best-effort slice. The resources in LTE are equally divided among the data users. However, the data rate of individual users in LTE is also limited by the policy in the network (D_A, say). This assumption is inline with the restrictions posed by UE Aggregate Maximum Bit Rate (AMBR) [3]. If there are j data users in LTE, let the total throughput obtained in LTE be denoted by $D_L(j)$. Therefore, $D_L(j) = \min\{jD_A, C_L\}$. The throughput obtained in LTE increases linearly with the number of users. However, it is limited by C_L. In WLAN, we consider the saturation throughput model [16] for data users. This model accurately characterizes the maximum throughput of the system. It is calculated based on the packet length, channel idle time and takes into account, the contention amongst users. If there are k users in WLAN, let the per-user throughput (in Mbps) be denoted by $D_W(k)$. The simulation parameters for WLAN and LTE RATs are described in Table 1 and have been obtained from [17]. The data user arrivals are assumed to follow Poisson processes and service times are exponentially distributed. We consider $D_A = 5$ Mbps and $C_L = 50$ Mbps.

Remark 1: We have used LTE and WLAN RATs for simulations as to the best of our knowledge, an open-source simulator for the 5G network is unavailable at present. Results for 5G NR RAT are expected to be similar (but scaled) to LTE due to the similarity in throughput behavior.

We validate the performance of the proposed architecture with the help of a simple association algorithm for a slice supporting best-effort data users. The optimal solution for user-association which maximizes the total system throughput for the best-effort slice can be obtained using the well-known value iteration algorithm [18] under suitable assumptions on the arrival process and service time of data users. However, it is known to have an exponential worst case computational complexity. This motivates us to propose a simple and greedy solution which has a polynomial computational complexity. Also, the proposed algorithm does not require the knowledge of the statistics of the distribution of user arrivals and hence, can be implemented in real time.

B. Description of a Simple User Association Algorithm

As described in Algorithm 1, whenever a user sends an association request to the multi-RAT controller via the LTE dBS/WLAN dBS, the multi-RAT controller looks at the number of active best-effort users in LTE and WLAN, viz., j and

[1] The source code for the evaluation platform is available at: https://infonetsdn@bitbucket.org/infonetsdn/multirat_sdn.git

TABLE 1: Multi-RAT network model for LTE and WLAN.

Parameter	**Value**
Mean service time for user	60s
Path loss	$128.1 + 37.6\log(R)$, R in kms
WLAN channel bit rate	54 Mbps
Tx power for LTE dBS	46 dBm
Tx power for WLAN dBS	23 dBm
Tx power for UE	23 dBm
Antenna Type (LTE and WLAN)	Isotropic Antenna

k. To evaluate the preferred RAT for the association, we need to evaluate the following boolean variable B.

$$B = I_{\{D_L(j+1)+kD_W(k)>D_L(j)+(k+1)D_W(k+1)\}}, \tag{1}$$

where $I_{\{.\}}$ denotes the indicator variable. After evaluating B using Equation (1), if we observe that $B = 1$, then the user is associated with LTE, else WLAN is selected. The association which provides a better throughput is chosen based on the current load of both LTE and WLAN dBSs. For example, consider that $D_A = 5$ Mbps, $C_L = 50$ Mbps, and we consider 802.11g [19] WLAN dBS. Calculation [16] reveals that $D_W(1) > D_A$. Therefore, when the system is empty, it is better to associate an incoming user with a WLAN dBS. However, the greedy scheme dictates that if before the departure of the associated WLAN user, another user arrives, it is better to associate the user with LTE since $2D_W(2) - D_W(1) < D_A$ and so on.

Algorithm 1 User association algorithm for best-effort slice in the proposed multi-RAT architecture.

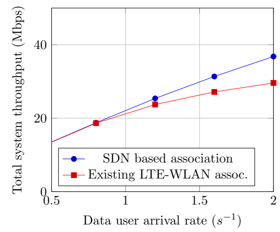
Initialize $j \leftarrow 0$, $k \leftarrow 0$ and $B \leftarrow 0$.
procedure USER–ASSOCIATION
  for each arrival of data users **do**
    Evaluate B using Equation (1).
    if $B == 1$ **then**
      Associate user with LTE dBS.
      $k \leftarrow k + 1$.
    else
      Associate user with WLAN dBS.
      $j \leftarrow j + 1$.
    end if
  end for
end procedure

Remark 2: The throughput behavior as a function of load of the network in 5G is analogous to that of LTE. Therefore, the proposed association scheme can be adopted without any modification in the case of a 5G dBS and a WLAN dBS.

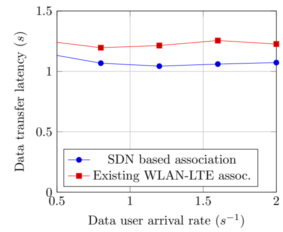
Remark 3: The proposed algorithm is used only as an illustrative tool to evaluate the capabilities of the architecture. However, more sophisticated approaches can be devised in future.

C. Simulation Results

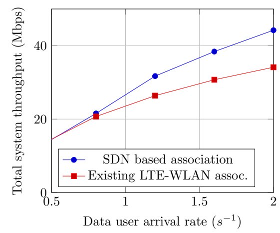
We measure the system throughput and end-to-end data transfer latency for different types of data traffic in both the networks. In existing multi-RAT networks consisting of overlapping LTE and WLAN coverage areas, an incoming user is always associated with WLAN until the WLAN AP denies association. All incoming users are then associated with the LTE eNB. We implement Algorithm 1 for RAT selection in the proposed network architecture. The performance improvements provided by our architecture vis-a-vis existing network architectures are demonstrated using three scenarios.



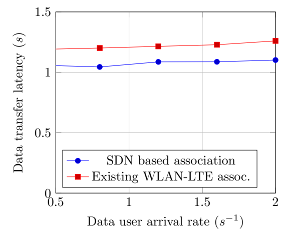
(a) System throughput v/s best-effort data user arrival rate (Scenario 1).



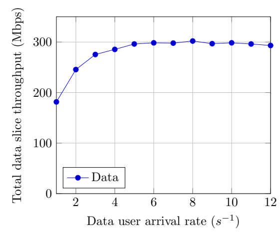
(b) Data transfer latency v/s best-effort data user arrival rate (Scenario 1).



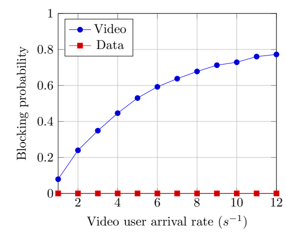
(c) System throughput v/s CBR data user arrival rate (Scenario 2).



(d) Data transfer latency v/s CBR user arrival rate (Scenario 2).



(e) Data slice throughput v/s data user arrival rate (Scenario 3).



(f) Video user blocking probability v/s data user arrival rate (Scenario 3).

Fig. 5: Simulation Results

1) Scenario 1 - Network slice supporting best effort traffic: As illustrated in Fig. 5a, the system throughput for the SDN based multi-RAT network is consistently better than that of existing networks for best effort data traffic. A performance improvement of $0.3\% - 24\%$ is obtained by implementing Algorithm 1 in the SDN based multi-RAT network. This is due to the fact that in the existing multi-RAT networks, a given RAT may not possess the load information of other RATs. In the proposed multi-RAT architecture, the presence of load information of all the constituent RATs at the controller improves user association decisions, leading to an improvement in the total system throughput. We also observe that the end-to-end data packet latency as illustrated in Fig. 5b is lower with increased arrival rate in the SDN framework in comparison to the existing network. The latency reduction is observed to be between $8\% - 12\%$. This is because within the present day networks, users are associated with WLAN until the capacity is reached, irrespective of load on the LTE RAT. This results in increased packet transfer delays and reduced system throughput as the contention in WLAN increases with increased user arrival rate.

2) Scenario 2 - Network slice supporting Constant Bit Rate (CBR) traffic : In this scenario, we evaluate the system throughput and the data transfer latency for a network slice that supports traffic requiring CBR for each user such as video, Voice over Internet Protocol (VoIP). In existing networks, if the LTE RAT is unable to provide QoS guarantees for such traffic requests due to lack of capacity, the requests are blocked. Within the SDN based multi-RAT network, we assume a load threshold configured as per the network orchestration policy for LTE and block all users after the threshold. We also consider a threshold on the number of WLAN data users to guarantee that the per-user throughput is always above a certain data rate. The load threshold for WLAN is calculated from [16]. As a result, we now block users in WLAN and LTE whenever their respective load thresholds are reached. Below the load thresholds, user associations are handled as specified in Algorithm 1. For this scenario, we assume that the constant bit rate for a user in LTE is 3 Mbps. The throughput for the proposed network architecture as illustrated in Fig. 5c is better for CBR traffic. Similarly, the latency for data transfer as shown in Fig. 5d is also lower in comparison to that of the existing network. This is because the approach followed by the proposed algorithm always associates the user with the best RAT, given the current load conditions of the system.

3) Scenario 3 - Isolation of Network Slices: In the third scenario, we illustrate that our framework supports network slicing and demonstrate that slice isolation can be ensured over a common network infrastructure. We consider the same system model as in the second scenario. The network is divided into two slices, one serving real-time video traffic (video-slice) and the other serving best-effort data traffic (data-slice). We envision that real time video traffic can be served by LTE RAT whereas best-effort data traffic can be served by either LTE or WLAN. Accordingly, the video slice consists of the LTE dBS, and data slice consists of both LTE and WLAN dBSs. The data rate for real-time video users are configured to be 400kbps. We implement a policy reserving 30% of the LTE resources for real-time video, and the rest are reserved for data. The WLAN dBS is reserved for data traffic.

We measure the system throughput of the best-effort data-slice by varying the arrival rate of data users and maintaining a constant video user arrival rate. As shown in Fig. 5e, the throughput of the data traffic increases upto the slice capacity and then remains constant. However, the throughput of the video slice remains unaffected. Similar observation holds when the arrival rate of video users is varied by keeping data user arrival rate constant. Video traffic requests that arrive after the slice capacity is reached, are blocked without affecting the traffic in any other slice. This is illustrated in Fig. 5f by the observed increase in blocking probability of video traffic with the growth in video user arrival rate.

V. Conclusions

In this paper, we have proposed an SDN based network architecture for unified control and management of multi-RAT networks. The architecture provides end-to-end network control while ensuring scalability through the creation of multiple logical networks or slices over a single physical network. It provides RAT-agnostic interfaces to applications and a virtualized view of the network resources, enabling simplified control and management. Additionally, we have demonstrated the performance improvements provided by this architecture through call flows and experimental results.

VI. Acknowledgement

The authors thank Ashish Sharma, Rohan Kharade and Abhishek Dandekar for their contributions towards the development of the SDN based evaluation platform in ns-3. This work has been supported by the Department of Telecommunications, Ministry of Communications, India as part of the indigenous 5G Test Bed project.

References

[1] Broadband Forum TR-321, "Public Wi-Fi Access in Multi-service Broadband Networks," 2015.

[2] A. Roy, P. Chaporkar, and A. Karandikar, "Optimal Radio Access Technology Aelection Algorithm for LTE-WiFi Network," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 6446–6460, 2018.

[3] 3GPP TS 23.501, "System Architecture for the 5G System," 2017.

[4] E. Haleplidis, K. Pentikousis, S. Denazis, J. H. Salim, D. Meyer, and O. Koufopavlou, "RFC 7426: Software-Defined Networking (SDN): Layers and Architecture Terminology," 2015.

[5] I. F. Akyildiz, P. Wang, and S.-C. Lin, "SoftAir: A Software Defined Networking Architecture for 5G Wireless Systems," *Elsevier Computer Networks*, vol. 85, pp. 1–18, 2015.

[6] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, "Network Slicing based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[7] M. M. Rahman, C. Despins, and S. Affes, "HetNet Cloud: Leveraging SDN and Cloud Computing for Wireless Access Virtualization," in *in IEEE International Conference on Ubiquitous Wireless Broadband*, pp. 1–5, 2015.

[8] K. S. Munasinghe, I. Elgendi, A. Jamalipour, and D. Sharma, "Traffic Offloading 3-tiered SDN Architecture for DenseNets," *IEEE Network*, vol. 31, no. 3, pp. 56–62, 2017.

[9] X. Mi, Z. Tian, X. Xu, M. Zhao, and J. Wang, "NO Stack: A SDN-based Framework for Future Cellular Networks," in *Wireless Personal Multimedia Communications*, pp. 497–502, 2014.

[10] Y. Fengyi, W. Haining, M. Chengli, Z. Jianmin, and W. Min, "A Flexible Three Clouds 5G Mobile Network Architecture based on NFV & SDN," *IEEE China Communications*, vol. 12, pp. 121–131, 2015.

[11] E. J. Kitindi, S. Fu, Y. Jia, A. Kabir, and Y. Wang, "Wireless Network Virtualization With SDN and C-RAN for 5G Networks: Requirements, Opportunities, and Challenges," *IEEE Access*, vol. 5, pp. 19099–19115, 2017.

[12] A. Nayak M, P. Jha, and A. Karandikar, "A Centralized SDN Architecture for the 5G Cellular Network," *in IEEE 5G World Forum*, pp. 147–152, 2018.

[13] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega, D. Aziz, and H. Bakker, "Network Slicing to enable Scalability and Flexibility in 5G Mobile Networks," *IEEE Communications magazine*, vol. 55, no. 5, pp. 72–79, 2017.

[14] R. Ferrus, O. Sallent, J. Perez-Romero, and R. Agusti, "On 5G Radio Access Network Slicing: Radio Interface Protocol Features and Configuration," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 184–192, 2018.

[15] 3GPP TS 29.244, "Interface between the Control Plane and the User Plane nodes," 2018.

[16] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, 2000.

[17] 3GPP TR 36.839, "Mobility Enhancements in Heterogeneous Networks," 2012.

[18] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[19] IEEE Computer Society LAN/MAN Standards Committee, "802.11-2016 Part I1: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 2016.