IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 18, 2020, accepted May 1, 2020, date of publication May 6, 2020, date of current version May 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992805

Dual Connectivity-Based Mobility Management and Data Split Mechanism in 4G/5G Cellular Networks

TARIQ MUMTAZ[1,2], SHAHABUDDIN MUHAMMAD[3], MUHAMMAD IMRAN ASLAM[1], AND NAZEERUDDIN MOHAMMAD[4]

[1]Department of Electronics Engineering, NED University of Engineering and Technology, Karachi 75270, Pakistan
[2]Dhanani School of Science and Engineering, Habib University, Karachi 75290, Pakistan
[3]Department of Computer Science, Prince Mohammad Bin Fahd University, Al-Khobar 31952, Saudi Arabia
[4]Cybersecurity Center, Prince Mohammad Bin Fahd University, Al-Khobar 31952, Saudi Arabia

Corresponding author: Tariq Mumtaz (tariqm@neduet.edu.pk)

This work was supported by Prince Mohammad Bin Fahd University (PMU).

ABSTRACT The emerging 5G mobile network technology is envisioned to provide an efficient platform to interconnect machines, objects, and devices in addition to interconnecting people. Equipped with peak data rates, low latency, and massive capacity, 5G technology will empower new user experiences such as virtual reality and augmented reality and provide new service areas such as connecting massive IoT. Dual connectivity is an important feature where 5G systems are overlaid on the existing 4G core network. In this paper, we propose an MM (mobility management) algorithm to efficiently perform handovers between 4G and 5G RATs (radio access technologies). Our proposed MM algorithm utilizes the strength of DC (dual connectivity) for MM as DC inherently has lesser amount of handover interruption as compared to conventional hard handover. Our MM scheme suggests appropriate data split mechanism between 4G and 5G RATs based on application-specific strategy. We provide a framework based on probabilistic model checking that leverages DC and suggests strategy-based data split mechanism for a mobile user for a variety of market verticals. We model the system as MDP (Markov decision process) where a controller breaks all the nondeterminism in the MDP based on reward calculations. The proposed framework is implemented in a well known model checker and various scenarios are used to assess its applicability.

INDEX TERMS 5G, 4G, mobility management, dual connectivity, split architecture, probabilistic model checking.

I. INTRODUCTION

The emerging wireless technologies will enable new differentiated services and applications to support the needs of the 4th industrial revolution. The services, such as the security service in a smart city through IoT (Internet of Things) [1] and virtual reality health care system [2], are expected to generate a huge volume of data that demands sharp capacity increase in wireless networks. Moreover, different applications in emerging wireless networks bring a diverse QoS (Quality of Service) requirements such as high data rate, low latency and highly reliable communication to meet desired user experience. To address the diversified requirements from different market verticals shown in Fig. 1, the 5G (fifth-generation) of mobile networks has been introduced. The goal of 5G is to provide flexible and scalable connectivity quickly and efficiently through reliable and low latency communication channels.

The targeted key enhancement areas of upcoming 5G networks include area traffic capacity, user experienced data rate, spectral efficiency, connection density, latency, and energy efficiency [3]. In the first phase of 5G deployment, the major focus is on capacity enhancement, increased data rate and spectral efficiency. 5G networks are expected to provide 10 to 100 times higher user data rate and cater 1000 times higher data volume in a given area as compared to current legacy networks. To meet the demand for massive capacity enhancement, each cell site of 5G network must have high

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenhui Yuan.

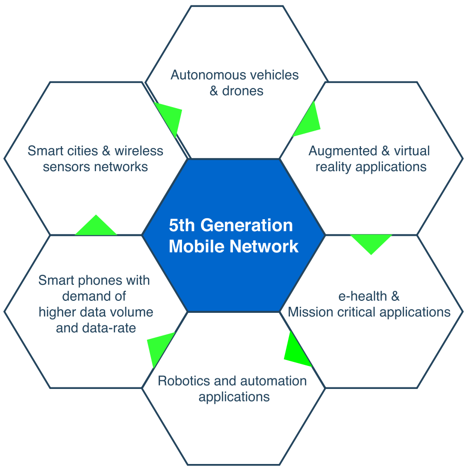


FIGURE 1. Wireless market verticals for 5G technology.

capacity available and the number of cell sites in a given area must be higher than legacy networks. Mobile operators are expecting huge deployment of new radio BSs (base stations) to provide improved coverage, enhanced capacity, and reliable radio communication link between a BS and a UE (user equipment).

Increasing the number of BSs reduces the average coverage footprint of BSs which results in the increased number of HOs (handovers) as compared to legacy networks. Increased number of HOs causes increase in signaling overhead during the HO process. This also results in frequent data transfer interruption associated with each HO procedure. So there is a trade-off between increased capacity through BS densification and the corresponding increase in signaling load and data interruption due to HOs. Therefore, 5G networks must have intelligent algorithms for MM (mobility management) to reduce the amount of control signaling and frequent interruptions by reducing the total number of HOs.

In this paper, we propose an algorithm to efficiently perform HO between 4G and 5G RATs (radio access technologies). We consider a two-layer network consisting of an existing 4G layer providing blanket coverage and a newly deployed 5G layer. The presence of the additional 5G layer is to provide high capacity, increased throughput, and improved spectral efficiency. Our framework consists of an algorithm that considers a variety of parameters such as cell density, coverage footprint, number of users, and radio channel state, to make HO decisions within the same layer (intra-RAT) or across different layers (inter-RAT).

For fast-moving users, the number of HOs may increase drastically due to frequent selection of small cells. This has a negative influence on user experience due to increased interruptions during data transfer. This situation is exacerbated in areas where 5G BSs are densely deployed. DC (dual connectivity) is an important feature where 5G systems are overlaid on the existing 4G core network. In DC configuration, users maintain connections to 4G as well as to 5G RATs in order to reduce interruptions in data transfers during HOs. Our proposed MM algorithm utilizes the strength of DC for MM, as DC inherently has lesser amount of HO interruption as compared to conventional hard HO. The proposed MM scheme also suggests appropriate data split mechanism between 4G and 5G RATs based on application-specific strategy.

The proposed framework is based on PMC (probabilistic model checking) that leverages DC and suggests application-specific data split mechanism for a mobile user for a variety of market verticals. We model the system as an MDP (Markov decision process) where a controller breaks all the nondeterminism in the MDP based on reward calculations. The proposed framework considers several key parameters based on network topology, RF (radio frequency), and discount factors to calculate the reward that guides the appropriate RAT selection as well as the corresponding data split. Network topology parameters include network type, location type, coverage type, number of users, and time slot. The RF parameters include carrier frequency, bandwidth, channel type, and path loss. Discount related parameters include load type, RF channel type, and HO type. The proposed framework is implemented in a well-known model checker and various scenarios are used to assess its applicability. We highlight the key elements of our contribution in the following:

1) We develop a framework for multi-RAT dual connectivity MM with split data mechanism. The framework models the system using MDP where a reward-based controller, represented as DTMC (discrete time Markov chain), guides the handover policy.
2) Instead of only considering RSS (received signal strength) of the target cell, the proposed model considers several key parameters to decide target cell selection for HO. In particular, it considers parameters related to mobile network topology as well as RF dependent parameters in reward calculation that guides the HO policy for target RAT selection.
3) To control the total number of HOs, a reward function is used that employs various discount factors to capture discounted rewards corresponding to RF channel state, BS load condition, and HO type. We incorporate split bearer architecture in MM mechanism to cater the impact of high number of HOs due to user velocity and 5G small cells density. The split radio bearer functionality routes data packets through alternative path for 5G users in DC configuration.
4) We present the comparison of our DC-based MM and data split mechanism with the state-of-the-art scheme FHM (frequent handover mitigation) for ultra-dense heterogeneous networks. It is shown that the proposed approach is very effective in maintaining good network throughput in high user velocity and high small cell density scenarios.
5) It is shown that the proposed split traffic mechanism gracefully reduces HO count at 5G layer based on various strategies.

6) The proposed framework serves as a guiding tool to mobile operators, vendors, and researchers to make application-specific judgment and select appropriate system parameters and data split strategy to attain the desired QoS.

The rest of the paper is organized as follows. Section II reviews the background related to 5G MM schemes. Section III introduces the DC configuration in 5G networks and describes the model, related parameters, and the proposed methodology. The important scenarios and the corresponding results are presented in section IV. Section V concludes the paper and presents possible scope for future work.

II. MOBILITY MANAGEMENT IN 5G

MM in cellular networks is an important research area that has received considerable attention over the past few years, specifically in the case of 5G networks where the level of system complexity is very high due to several applications in diverse market verticals. HO process of MM plays a very critical role in maintaining call continuity and ensuring targeted QoS for mobile users.

The HO process comprises of three phases: preparation phase, execution phase, and completion phase. The HO preparation phase is an important stage where certain criteria based on different parameters need to be met in order to trigger HO. Our work focuses at this phase of HO. In this phase, the decision is made to select the appropriate target cell and the system decides the type of HO (such as intra-cell, inter-frequency, or inter-RAT).

In [4], Akshay *et al.* present a detailed design and implementation challenges of MM to deal with highly dense heterogeneous 5G networks. The authors posit that the existing MM schemes are not able to meet the requirements of flexibility, scalability, and reliability to ensure strict 5G QoS targets. They identify challenges and potential MM solutions for 5G which include SDN (software define networks) based smart core network signaling, on demand and adaptive MM schemes to serve users with different mobility profiles for diverse market verticals, and efficient utilization of resources of the network with different load dynamics.

In [5], Yun Li *et al.* present MM survey related to different HO techniques like RSS (received signal strength) based HO. In this type of HO, the UE is directed by the BS to HO to a cell that has the highest RSS to ensure best possible communication channel between a UE and a BS. Although RSS-based HO scheme is simple and fast, it does not take other parameters into consideration such as resource utilization and bandwidth available in the target cells. There might be instances where a highly utilized cell with low bandwidth gets selected as the target cell for the HO, which can degrade the perceived QoS for users.

In [6], Angela *et al.* propose mobility aware user association for 5G mmWave networks which can cater limitation of conventional RSS-based cell selection approach. Their proposed model incorporates additional factors during cell selection such as changing RF conditions due to user mobility, BS load, and mmWave related aspects which include non-line-of-sight propagation and sensitivity to blockages. Through Monte Carlo simulation, it is shown that the proposed association is significantly beneficial and the polynomial-time algorithm is computationally faster than exhaustive search algorithm.

In [7], Arshad *et al.* present HO management in 5G using topology aware skipping approach. Their work presents the challenge of 5G network densification in the context of increased number of HOs and the associated delays which leads to throughput degradation. To avoid excessive HOs, the authors propose an HO scheme which takes user location and cell size into consideration when deciding to skip a cell for HO. The proposed scheme results in reduced HO cost and increased user throughput as compared to other approaches such as best connected and alternating skipping. Their work also highlights the impact of interference caused by skipped BS on user perceived SINR (signal to interference plus noise ratio). To address this issue, authors suggest to incorporate IC (interference cancellation) and CoMP (coordinated multi point transmission).

In [8], Wang *et al.* propose RL (reinforcement learning) two-layers HO control-based scheme for UDNs (ultra dense networks) with heterogeneous user mobility patterns. In their work, the authors present the idea of clustering UEs that have similar mobility patterns. Synchronous deep reinforcement learning is applied to develop optimal HO management scheme for UEs. The authors demonstrate the performance of the proposed scheme in terms of reduced HO count and improved throughput. Their work addresses the problem of system performance degradation during early stages of learning for new arriving UEs by introducing SL (supervised learning) techniques.

In [9], Gharsallah *et al.* present a SDN and NFV (network function virtualization) based HO management approach for 5G UDN. The authors utilize all three planes of SDN architecture to design SDHO (SDN-based HO) scheme. The data plane comprises of high number of small cells with diverse technologies, the control plane consists of a controller for all centralized functions, and the application plane is used to create abstract view of network through NFV to determine the desired QoS. In the application layer, they define SDHME (software defined HO management engine) which is responsible for V-cell (virtual cell) creation through NFV. The V-Cell is a logical cell comprising of different small cells that serve a mobile user based on user's velocity and motion trajectory. V-Cell helps to expedite the HO process by predicting the next small cell, initiating resource activation in advance in the target cell, and releasing the resources from the departure cell. Through simulations, authors demonstrate that V-cell helps in reducing HO failure rate and HO delays.

The authors in [10] consider mmWave scenario in 5G where radio conditions are more challenging due to intermittent links. The cell selection process takes into account several access network parameters such as BS capacity, channel state conditions (*LOS*: line-of-sight, *NLOS*: non-line-of-sight, and *outage*), number of attached users in the target BS, and most importantly signaling overhead which occurs during HO execution phase. Their work utilizes PMC to design mobile network controller and evaluates different MM schemes for target cell selection during HOs. The authors propose an efficient MM algorithm that enhances the network utilization and improves user perception by selecting better channel type.

In [11], Hasan *et al.* describe a scenario of excessive HOs in LTE ultra dense small cell deployment for fast moving users and frequent ping-pong HOs for slow moving or static users due to varying radio conditions like channel fading and shadowing. The proposed scheme for undesired HO mitigation works in two phases. In the first phase, unwanted HOs are classified as fast moving or ping-pong. In the second phase, ping-pong HOs are mitigated through soft parameters such as HO hysteresis which can stop HO execution. The fast moving HO category is treated through dual cell feature where the fast moving users are handed over to macro cell. The authors present the improvement in terms of reduced HO count and improved user throughput due to less packet transfer interruption.

In [12], Polese *et al.* propose a DC-based HO to cater intermittent RF channel states in mmWave 5G networks. The authors discuss specifically the challenge of RF outage state in mmWave and argue that typical hard HOs are not suitable for 5G users who demand ultra low latency specifically during HO process. In their research work, authors propose efficient, fast, and robust MM scheme based on DC. They implement 5G advance features like fast switching, secondary cell HO and adaptive TTT (time-to-trigger) to achieve less interruption during HO as compared to typical standalone hard HO. With the help of simulations, the authors demonstrate that their proposed MM algorithm is able to improve some QoS-driven KPIs (key performance indicators) like HO latency, throughput, and packet loss.

In [13], Antonioli *et al.* propose a DC-based dynamic split bearer for multi-RAT 5G networks. The authors propose flow control mechanism for user data split between two nodes of DC, that is, master and secondary nodes. The proposed flow control mechanism is adaptive to ensure the best split ratio for maintaining target QoS under varying RF channel conditions. On the basis of perceived QoS and RF channel quality received from UEs, the proposed algorithm targets utility-based resource allocation problem and maximizes the user satisfaction by suggesting appropriate utility weight to have better control on bearer split. The results show that the proposed bearer split-based flow control scheme improves user satisfaction level and throughput at both UE as well as system level.

Overall, the MM schemes presented above are able to handle the issue of high number of HOs to some extent, particularly in the planning phase of HO process. The presented techniques consider specific simulation scenarios which typically depend on complex and bounded implementation details in simulation tools. Formal verification techniques such as model checking, on the other hand, can capture the system under consideration at a variety of abstraction levels. Our main contribution is to present a formal model of the MM system and then apply PMC to assess the model against various dynamic system specifications for a variety of input parameters. To the best of our knowledge, this is the first time PMC is used to assess DC-based split configuration for MM in heterogeneous cellular networks. Our proposed framework considers a variety of parameters based on network topology and radio frequency to calculate the reward used in RAT selection. Additionally, it takes user velocity and 5G cell density into account to decide data split policy for various QoS requirements.

III. OUR WORK

This section describes the model, related parameters, and algorithms used in the proposed methodology. Since our work relies on DC split configuration in 5G networks to make efficient HO decision, we first briefly describe the basic architecture in the following subsection.

A. DC SPLIT CONFIGURATION IN 5G

Typical cellular network architecture is distributed into two main parts, the RAN (radio access network) and the core. The RAN mainly consists of radio BS and the transmission connectivity with the core of the network. To support different market verticals and meet strict QoS targets, split-RAN architecture is used in 5G where data and control planes work independently. The data and control plane split provides flexibility of deployment topology in RAN architecture.

Due to split architecture, two possible options are selected and standardized by 3GPP for BS connectivity with the core network in 5G [14]. First option is SA (standalone architecture) where 5G radio node (called gNB) is connected with 5G core. The other option is NSA (non-standalone architecture) where gNB is connected to 4G core. This approach permits network operators to roll out 5G services to market in shorter time without the expense of a full scale 5G core network. The gNB connectivity with 4G core is shown in Fig. 2.

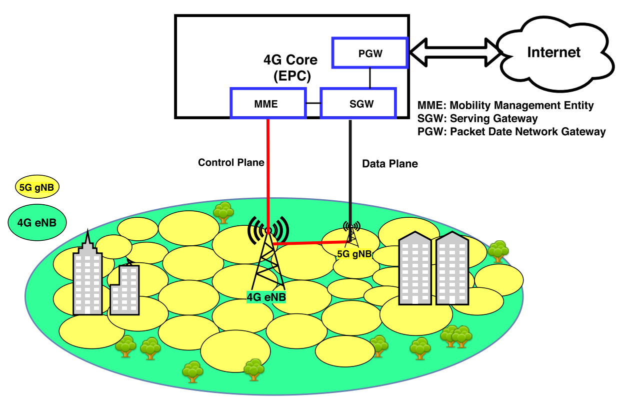


FIGURE 2. Non standalone configuration (NSA).

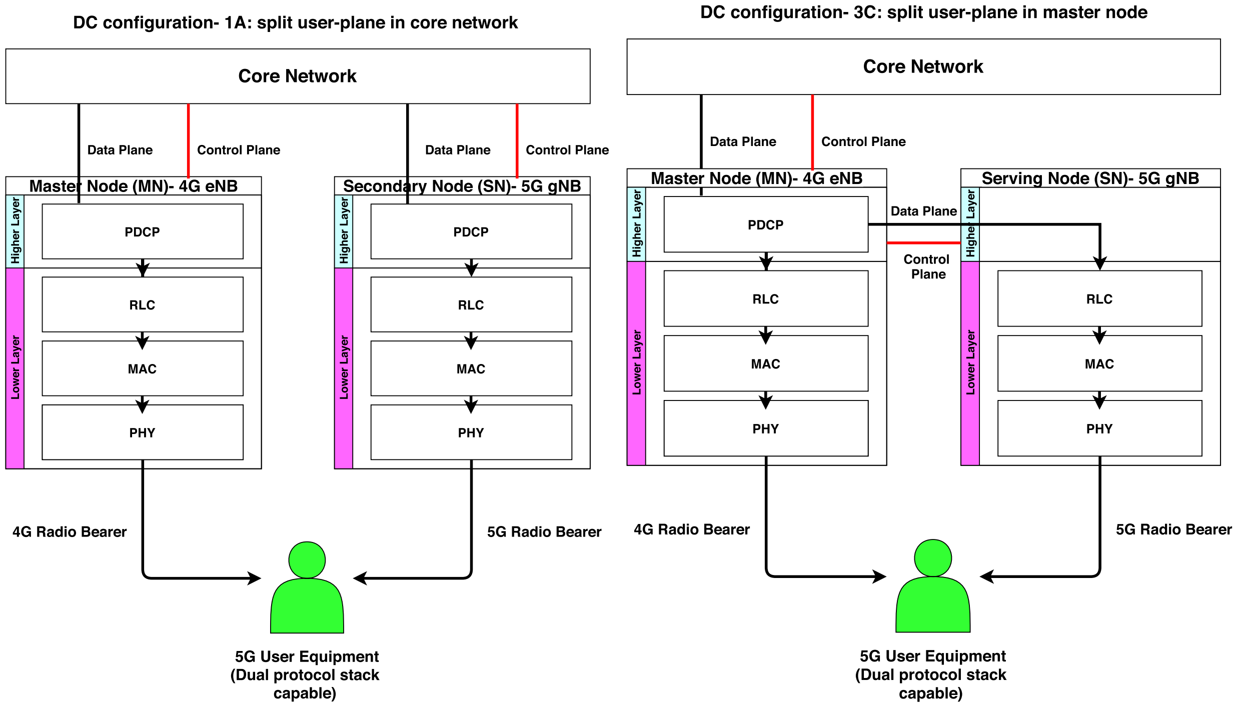


FIGURE 3. DC user plane configuration.

The DC feature tightly couples 4G and 5G technologies in NSA [14]. In DC, a UE gets connectivity from two cells simultaneously. These cells are called MN (master node) and SN (secondary node). The MN provides both control and user plane connectivity to a user while SN provides only user plane connectivity. The MN is responsible for all control plane signaling such as connection setup, connection release and HO management (like SN node addition and SN change or switching). After the successful establishment of control plane, user data transfer starts through user plane using DRB (data radio bearer). The gNB user plane protocol stack consists of higher layers which include SDAP (service data adaptation protocol) and PDCP (packet data convergence protocol) whereas lower layers of user plane include RLC (radio link control), MAC (medium access control), and PHY (physical) layers.

The DRB in 5G is broken into two categories, namely direct and split data bearer [15]. In direct DRB, the user plane of higher layers uses lower layers of the same node for data packet transfer. In split DRB, higher layers can use lower layers of other nodes for packet transfer. This split bearer is an important aspect of 5G architecture and is considered vital for meeting strict targets such as latency and reliability. Currently 3GPP standard has two configurations of user plane, 1A configuration where user data split occurs at the core network, and 3C configuration where user data split occurs at MN. DC user plane dual configuration is shown in Fig. 3. The set of notations related to 4G and 5G technologies is summarized in Table 1.

TABLE 1. Notations related to 4G and 5G communication technologies.

Notation	Meaning
BS	Base station
UE	User equipment
MM	Mobility management
RAT	Radio access technology
HO	Handover
RSS	Received signal strength
RF	Radio frequency
IC	Interference cancellation
SINR	Signal to noise ratio
CoMP	Coordinated multi point
UDN	Ultra dense network
RAN	Radio access network
gNB	Logical 5G radio node
eNB	4G-LTE evolved node B
SA	Standalone architecture
NSA	Non standalone architecture
DC	Dual connectivity
MN	Master node
SN	Secondary node
DRB	Data radio bearer
SDAP	Service data adaptation protocol
PDPC	Packet data convergence protocol
RLC	Radio link control
MAC	Medium access control

B. THE PROPOSED MODEL

We consider a two-layer heterogeneous network. The first 4G layer provides a complete blanket coverage while the second layer consists of 5G cells with short coverage footprints. We assume urban topology where cell sites are uniformly placed to provide RF coverage to 5G users of diverse verticals. For the sake of simplicity, the coverage region of each

TABLE 2. Definition of frequency ranges.

Frequency range designation	Corresponding frequency range
FR1	410 MHz - 7125 MHz
FR2	24250 MHz - 52600 MHz

cell is assumed circular. As there are many small 5G cells under the larger coverage area of a 4G cell, we propose a smart MM algorithm based on NSA configuration with DC feature. The larger footprint 4G cell serves as MN while 5G small cells serve as SNs. As discussed earlier, all the signaling towards 4G core on the control plane is controlled by 4G cell while user data packet transfer is through high capacity 5G cells.

To cater high capacity demands in 5G, two frequency bands are considered for 5G deployment as shown in Table 2 [16]. FR1 frequency band has been selected for 5G phase-I where each cell has a maximum 100 MHz bandwidth capacity. We consider FR1 for our proposed MM model as well. For the 4G layer, the carrier frequency of 2 GHz is used in our system model with 20 MHz bandwidth. We assume RF coverage criteria as a prerequisite and the process starts only if the source and target cells meet coverage thresholds as per the mobile network operator's strategy.

To cater the diverse set of mobility scenarios for different market verticals under varying radio access layer conditions, we model the system as MDP (referred to as M_1). Further, the controller (referred to as M_2) is modeled as DTMC which helps in HO decision making. The system M_1 sends relevant information to the controller M_2 in every *epoch*. Model checking uses an abstract notion of time (called *epoch*) which captures relative ordering of state space during model execution. The controller calculates the reward and hence breaks the nondeterminism in the MDP by suggesting the desired HO policy. PMC is used to verify a variety of system specifications against the model. Interested readers may refer to [17] for a comprehensive background on how to model stochastic processes (using MDP, DTMC, etc.), capture system specifications in temporal logic, and check the specification against the model using PMC. The details of the model and the associated parameters related to radio, network topology, and UEs are described in the following subsections.

1) UE VELOCITY

Our model considers user velocity v as one of the important parameters to capture a wide range of 5G user mobility profiles. 5G verticals include wide range of cases such as wireless sensors, indoor users (3 km/hr), urban vehicle (30 km/hr), open/rural vehicle (120 km/hr), and very high velocity trains (500 km/hr). A user's stay time in 5G cell is given in terms of the number of 5G slots as follows:

$$n^{5g}_{slots} = \begin{cases} \dfrac{\beta_{4g}}{v \cdot \rho_{5g}} & \text{if } \rho_{5g} > \gamma_{fp} \\ \dfrac{\beta_{5g}}{v} & \text{otherwise} \end{cases} \tag{1}$$

where ρ_{5G} is the density (that is, the number of BSs per unit area) of 5G cells, β_{4g} and β_{5g} are the footprints of 4G and 5G BSs respectively, and $\gamma_{fp} = \beta_{4g}/\beta_{5g}$ is the footprint ratio. At every time $epoch \in \mathbb{N}$, a user is at a distance d from the BS given by the following equation.

$$d = v \cdot \left| epoch - \left\lceil \frac{n^{rat}_{slots}}{2} \right\rceil \right| \tag{2}$$

This TX-RX distance d is an important parameter to capture RF channel states that a user receives at any time. It is useful in estimating whether the end user is in *LOS* or *NLOS* with the BS. Also, d is used in finding path loss to determine the RF coverage level at user's end.

2) 5G CHANNEL STATES

Various statistical models have been developed to predict channel state for different topologies like urban macro, rural macro, urban micro and in-house micro. These models help in finding whether a user is situated within clear *LOS* of the BS or resides in *NLOS* region due to some obstruction [18]. For 5G, we use 3GPP model to calculate *LOS* probability for UMa (urban macro) topology [19]. The inter-site distance for UMa model is not more than 500 m. The probability that a user experiences *LOS* using 3GPP UMa model is given as follows:

$$f_{los} = \min(c_1/d, 1) \cdot (1 - \exp(-d/c_2)) + \exp(-d/c_2) \tag{3}$$

where c_1 and c_2 are curve fitting parameters and their values are 18 and 63 respectively.

In addition to *LOS* and *NLOS* channel states, our model takes into account the situation where a user does not receive 5G signals. This channel state is named *outage*. The probability of outage is calculated as:

$$p^{5g}_{out} = \begin{cases} 1 - \dfrac{\rho_{5g}}{\gamma_{fp}} & \text{if } \rho_{5g} < \gamma_{fp} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Now, the probability that a channel is in *LOS* or *NLOS* state can be given in terms of outage probability and f_{los} as follows:

$$p^{5g}_{los} = (1 - p^{5g}_{out}) \cdot f_{los} \tag{5}$$

$$p^{5g}_{nlos} = 1 - p^{5g}_{los} \tag{6}$$

3) 4G CHANNEL STATES

The coverage area of 4G cell is wider in contrast to smaller 5G coverage areas targeted to cover specific regions. The 3GPP TR 38.901 RMa model suggests that the BS antenna height range from 10 m to 150 m and the inter-site distance go up to 5000 m [19]. This model is appropriate for 4G layer design in our system. The 4G *LOS* probability using RMa model is given as follows:

$$p^{4g}_{los} = \begin{cases} 1 & d \leq 10 \text{ m} \\ \exp(-\dfrac{d - 10}{1000}) & d > 10 \text{ m} \end{cases} \tag{7}$$

The corresponding 4G *NLOS* probability can now be given as:

$$p^{4g}_{nlos} = 1 - p^{4g}_{los} \tag{8}$$

4) RF PATH LOSS OF 5G SIGNAL

The user perceived RF signal level depends on user's distance from serving BS. Path loss is used to find the signal level at user's location. Several models have been used to calculate path loss for different topologies. The 3GPP UMa TR 38.901 path loss model [19] is a breakpoint model (d_{BP}) which takes BS height (h_{BS}) and UE height (h_{UE}) into consideration. But for cells with short coverage footprints like our 5G small cell model where coverage radius is in 100 to 300 m range, the breakpoint factor is not a necessary requirement. Hence, for small cells close-in (CI) free space reference distance 5GCM UMa model provides a similar path loss prediction which is much simpler equation. In fact, for CI model with 1 m reference distance 3GPP TR 38.901 UMa model transforms to 5GCM UMa model. Therefore, we consider 5GCM UMa path loss model [20] to assess the performance of 5G radio channel. It is frequency dependent and can be applied for carrier frequency range from 0.5 to 100 GHz band.

$$PL_c = 32.4 + 10nlog_{10}(d) + 20log_{10}(f_{GH}) + X^{Cl}_{\sigma} \tag{9}$$

Here f_{GH} is the carrier frequency in GHz, n is path loss coefficient, d is the distance in meters and X^{Cl}_{σ} is the shadow fading in dB. We use $f_{GH} = 6$, $n = 2$ and $X^{Cl}_{\sigma} = 4.1$ dB for $c = los$, and $n = 3$, $X^{Cl}_{\sigma} = 6.8$ dB for $c = nlos$.

5) RF PATH LOSS OF 4G SIGNAL

We consider 3GPP TR 38.901 RMa model [19] to represent path loss for 4G as given below:

$$PL_{los} = \begin{cases} PL_1 & 10m \leq d \leq d_{BP} \\ PL_2 & d_{BP} \leq d \leq 10\text{ km} \end{cases} \tag{10}$$

where PL_1 and PL_2 are given as:

$$\begin{aligned} PL_1 &= 20\log_{10}(40\pi d f_c/3) + \min(0.03h^{1.72}, 10)\log_{10}(d) \\ &\quad - \min(0.044h^{1.72}, 14.77) + 0.002\log_{10}(h)d \qquad (11) \\ PL_2 &= PL_1(d_{BP}) + 40\log_{10}(d/d_{BP}) \qquad (12) \end{aligned}$$

where h is the building height and d_{BP} is the break point distance.

$$d_{BP} = 2\pi \cdot h_{BS} \cdot h_{UE} \cdot f_c/c \tag{13}$$

where h_{BS} is the BS antenna height and h_{UE} is the UE antenna height in meters, f_c is the carrier frequency and c is the speed of light. The value of path loss for *nlos* channel type is given as:

$$PL_{nlos} = \max(PL_{los}, PL'_{nlos}) \tag{14}$$

where PL'_{nlos} is given as:

$$\begin{aligned} PL'_{nlos} &= 161.04 - 7.1\log_{10}(W) + 7.5\log_{10}(h) \\ &\quad -(24.37 - 3.7(h/h_{BS})^2)\log_{10}(h_{BS}) \\ &\quad +(43.42 - 3.1\log_{10}(h_{BS}))(\log_{10}(d) - 3) \\ &\quad +20\log_{10}(f_c) - (3.2(\log_{10}(11.75h_{UT})^2 - 4.97)) \end{aligned} \tag{15}$$

where W is the street width. We use $h_{BS} = 30$ m, $h_{UT} = 1$ m, $f_c = 2100$ MHz, $h = 5$ m and $W = 20$ m.

6) BS LOAD PROFILE

The number of UEs in specific RAT (4G or 5G) depends on the load profile of the BS. In general, we consider three load profiles as follows:

$$\text{load profile} = \begin{cases} N_{bs} & \text{high} \\ N_{bs} \cdot \delta_l & \text{low} \\ N_{bs} \cdot \delta_m & \text{medium} \end{cases} \tag{16}$$

where N_{bs} is the maximum number of UEs allowed to be attached to a BS of 4G RAT and δ_l and δ_m are the discount factors for light and medium load conditions respectively. Since the available capacity of 5G RAT is higher than 4G, 5G BSs are able to handle more users. We consider 5G area capacity factor c_a in our model to capture scaling factor for attached users in each 5G cell. Thus, the number of users in a 5G cell can be calculated as $N_{bs} \cdot c_a/\rho^{5g}$.

7) THE REWARD CALCULATION AT THE EDGE OF COVERAGE

When a UE reaches the edge of its current RAT's coverage, it needs to decide whether to stay on the same RAT causing intra-RAT HO, or to switch to the other RAT causing inter-RAT HO. The reward function for a 4G or 5G user depends on the available capacity in the target BS, spectral efficiency of the RAT, and corresponding BS load profile. The expression for the reward is given as follows:

$$r_{4g} = \begin{cases} \dfrac{\eta_1 \cdot BW_{4g}}{l^n_{4g}} \cdot (1 - \alpha_a) & \text{if } lt^c \in lt_{4g} \\ \dfrac{\eta_1 \cdot BW_{4g}}{l^n_{4g}} \cdot (1 - \alpha_i) & \text{otherwise} \end{cases} \tag{17}$$

where η_1 is the spectral efficiency in 4G, and α_a and α_i are overheads due to intra- and inter-HOs respectively. BW is the bandwidth, lt^c is the channel type that the user is currently attached to, and l^n_{4g} is the updated load of the target BS. The target for 5G peak spectral efficiency η_2 is 30 bits/Hz [21], which is 3 times higher than the existing 4G network spectral efficiency η_1. In our model, we use spectral efficiency scaling factor equal to η_2/η_1. The corresponding r_{5g} can be computed in the similar way.

Reward functions based on channel type in 4G and 5G RATs are given as:

$$\Gamma_{4g} = \begin{cases} r_{4g} \cdot \delta_{nlos} & \text{if } lt^n_{4g} = nlos \\ r_{4g} & \text{otherwise} \end{cases} \tag{18}$$

$$\Gamma_{5g} = \begin{cases} r_{5g} \cdot \delta_{nlos} & \text{if } lt^n_{5g} = nlos \\ r_{5g} \cdot \delta_{out} & \text{if } lt^n_{5g} = outage \\ r_{5g} & \text{otherwise} \end{cases} \tag{19}$$

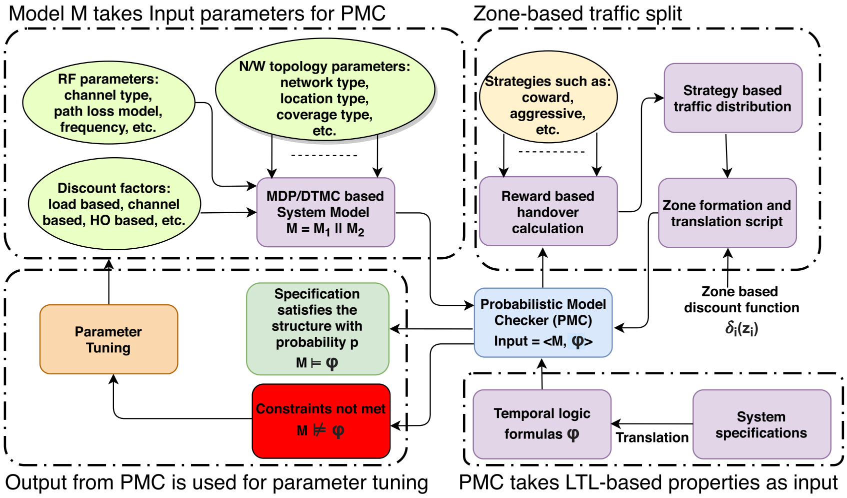


FIGURE 4. Overall working of the methodology.

where δ_{nlos} and δ_{out} are the discount factors due to *NLOS* and *outage* channel types respectively. δ_{nlos} can be calculated as follows:

$$\delta_{nlos} = \frac{|PL_{nlos} - PL_{los}|}{PL_{los}} \tag{20}$$

where PL_c is the path loss due to channel type $c \in \{los, nlos\}$.

8) THE REWARD CALCULATION FOR CHANGE IN CHANNEL TYPE

If a user is under the coverage of same cell within the duration of its stay time as described earlier, then the HO decision is made when the user observes any change in its current channel type. The new reward expression can be calculated as follows:

$$r_{stay} = \begin{cases} \dfrac{\eta_1 \cdot BW_{4g}}{l^c_{4g}} & \text{if } lt^c \in lt_{4g} \\ \dfrac{\eta_2 \cdot BW_{5g}}{l^c_{5g}} & \text{otherwise} \end{cases} \tag{21}$$

where l^c_{rat} is the load of the user's current BS. The corresponding reward based on channel types is given as:

$$\Gamma_{stay} = \begin{cases} r_{stay} \cdot \delta_{nlos} & \text{if } lt^c_{4g/5g} = nlos \\ r_{stay} \cdot \delta_{out} & \text{if } lt^c_{5g} = outage \\ r_{stay} & \text{otherwise} \end{cases} \tag{22}$$

On the other hand, if the user makes an HO, the corresponding reward calculations are given as follow:

$$r_{ho} = \begin{cases} \dfrac{\eta_2 \cdot BW_{5g}}{l^n_{5g}} \cdot (1 - \alpha_i) & \text{if } lt^c \in lt_{4g} \\ \dfrac{\eta_1 \cdot BW_{4g}}{l^n_{4g}} \cdot (1 - \alpha_i) & \text{otherwise} \end{cases} \tag{23}$$

$$\Gamma_{ho} = \begin{cases} r_{ho} \cdot \delta_{nlos} & \text{if } (lt^c \in lt_{4g} \wedge lt^n_{5g} = nlos) \\ & \vee (lt^c \in lt_{5g} \wedge lt^n_{4g} = nlos) \\ r_{ho} \cdot \delta_{out} & \text{if } lt^c \in lt_{4g} \wedge lt^n_{5g} = outage \\ r_{ho} & \text{otherwise} \end{cases} \tag{24}$$

Finally, the corresponding HO types (inter-HO ho_i, intra-HO ho_a) can be computed as:

$$\Omega = \begin{cases} ho_i & \text{if } (lt^c \in lt_{4g} \wedge \Gamma_{5g} > \Gamma_{4g}) \\ & \vee (lt^c \in lt_{5g} \wedge \Gamma_{4g} > \Gamma_{5g}) \\ & \vee (\Gamma_{ho} > \Gamma_{stay}) \\ ho_a & \text{if } (lt^c \in lt_{4g} \wedge \Gamma_{4g} > \Gamma_{5g}) \\ & \vee (lt^c \in lt_{5g} \wedge \Gamma_{5g} > \Gamma_{4g}) \end{cases} \tag{25}$$

That is, inter-RAT HO occurs when a users experiences higher reward either due to a newly encountered BS or due to changes in the current channel's condition. Otherwise, the user stays in the same BS either after making intra-RAT HO or by totally avoiding HO.

C. THE METHODOLOGY

The overall block diagram of the proposed methodology is shown in Fig. 4 where the model M, represented as parallel composition of system's model M_1 and controller's model M_2 (that is, $M = M_1||M_2$), takes RF parameters, discount factors, and network parameters as input. The algorithms describing the system M_1 as well as the controller M_2 are given in Algorithms 1 and 2, respectively. The following describes the main components of the system in further detail.

1) SYSTEM AND CONTROLLER MODELS

The algorithm of system's model M_1 begins by setting all the values of the relevant parameters (lines 2-3) such as load (current l^c and next l^n), link (or channel) type (current lt^c and next lt^n), and a boolean variable *cLTc* indicating whether the current link type has been changed. The system's model runs in an infinite loop (line 4) where it continuously waits for UEs

Algorithm 1 The System's Model M_1

1: **procedure** System M_1
2: Init: $\forall j \in BSs$:initialize $epoch,\ l^n_{4g},\ l^n_{5g},\ l^c,$
3: Init: $lt^c,\ lt^n_{4g},\ lt^n_{5g},\ cLTc$
4: **while** *true* **do** ▷ Continuously run for all epochs
5: $lt' = lt^c$
6: **if** $lt^c \in lt_{4g}$ **then**
7: $lt^c \leftarrow \{los, nlos\}$ with $p^{4g}_{los}, p^{4g}_{nlos}$
8: **else** $lt^c \leftarrow \{los, nlos, out\}$ with $p^{5g}_{los}, p^{5g}_{nlos}, p^{5g}_{out}$
9: **if** $lt^c \neq lt'$ **then** $cLTc = \top$
10: **else** $cLTc = \bot$
11: **if** L^c_{upd} **then**
12: $l^c_{4g} \leftarrow f_1(l^c_{4g_1}, l^c_{4g_2}, l^c_{4g_3})$
13: $l^c_{5g} \leftarrow f_1(l^c_{5g_1}, l^c_{5g_2}, l^c_{5g_3})$
14: **if** $\neg cLTc \wedge \neg RAT_{new} \wedge (epoch < T_e)$ **then**
15: $epoch \leftarrow epoch + 1$, **go to** 5
16: **if** $cLTc \wedge \neg RAT_{new}$ **then**
17: **if** $lt^c \in lt_{4g}$ **then**
18: $lt^n_{5g} \leftarrow \{los, nlos, out\}$ ▷ with corr. prob.
19: **else** $lt^n_{4g} \leftarrow \{los, nlos\}$ ▷ with corr. prob.
20: **if** RAT^{4g}_{new} **then**
21: $lt^n_{4g} \leftarrow \{los, nlos\}$ ▷ with corr. prob.
22: $l^n_{4g} \leftarrow f_2(l^n_{4g_1}, l^n_{4g_2}, l^n_{4g_3})$
23: **if** RAT^{5g}_{new} **then**
24: $lt^n_{5g} \leftarrow \{los, nlos, out\}$ ▷ with corr. prob.
25: $l^n_{5g} \leftarrow f_2(l^n_{5g_1}, l^n_{5g_2}, l^n_{5g_3})$
26: **if** RAT^{all}_{new} **then**
27: $lt^n_{4g} \leftarrow \{los, nlos\}$ with corr. prob.
28: $l^n_{4g} \leftarrow f_2(l^n_{4g_1}, l^n_{4g_2}, l^n_{4g_3})$
29: $lt^n_{5g} \leftarrow \{los, nlos, out\}$ with corr. prob.
30: $l^n_{5g} \leftarrow f_2(l^n_{5g_1}, l^n_{5g_2}, l^n_{5g_3})$
31: [*Sync*] with *Controller*
32: **if** $\Omega = ho_i \wedge lt^c \in lt_{4g}$ **then** $lt^c \leftarrow lt^n_{5g}$
33: **if** $\Omega = ho_a \wedge lt^c \in lt_{4g}$ **then** $lt^c \leftarrow lt^n_{4g}$
34: **if** $\Omega = ho_i \wedge lt^c \in lt_{5g}$ **then** $lt^c \leftarrow lt^n_{4g}$
35: **if** $\Omega = ho_a \wedge lt^c \in lt_{5g}$ **then** $lt^c \leftarrow lt^n_{5g}$
36: $epoch \leftarrow epoch + 1$

to guide their HO policy based on controller's feedback. lt^c is set based on Eq. 4 to Eq. 8 (lines 6-8). $cLTc$ is set true ($\top$) or false ($\bot$) based on whether there is any change in the channel type (lines 9-10). Current load profile is updated (lines 12-13) based on Eq. 16. The while loop continues for the next epoch (within the threshold T_e) if there is no change in the channel or no new RAT has been observed (lines 14-15). On the other hand, the channel type lt^n for both RATs (and the corresponding load if new RAT is observed) are updated (lines 16-30). Finally, the system synchronizes with the controller (line 31 in M_1 and line 3 in M_2) which breaks the nondeterminism in M_1 based on reward calculations.

Algorithm 2 The Controller's Model M_2

1: **procedure** Controller M_2
2: Init: $ho_i = \bot,\ ho_a = \bot$ ▷ $\top$ = true, $\bot$ = false
3: [*Sync*] with *System*
4: **if** $\neg RAT_{new}$ **then**
5: **if** $\Gamma_{ho} > \Gamma_{stay}$ **then** ▷ ref. eq. (22), (24)
6: $ho_i \leftarrow \top \wedge ho_a \leftarrow \bot$
7: **else**
8: $ho_i \leftarrow \bot \wedge ho_a \leftarrow \bot$
9: **else** ▷ ref. eq. (18), (19)
10: **if** $((lt^c \in lt_{4g}) \wedge (\Gamma_{5g} > \Gamma_{4g}))\vee$
11: $((lt^c \in lt_{5g}) \wedge (\Gamma_{4g} > \Gamma_{5g}))$ **then**
12: $ho_i \leftarrow \top \wedge ho_a \leftarrow \bot$
13: **else**
14: $ho_i \leftarrow \bot \wedge ho_a \leftarrow \top$

The controller's model M_2 in Algorithm 2 calculates the corresponding reward based on the input from M_1, that is, Γ_{ho} or Γ_{stay} if no new RAT is observed (lines 4-8) or Γ_{4g} or Γ_{5g} otherwise (lines 9-14). The reward values are used by M_1 to make HO decision (lines 32-35).

2) PRISM IMPLEMENTATION

We use PRISM, one of the widely used probabilistic model checker [22], that takes the model M and the system's specifications (translated as temporal logic formulas φ) as input and verifies the model against φ. PRISM uses reward structures to capture rewards specific to any state (or state-transition) in the model. For example, the following reward structure is used to calculate the number of times a user makes inter-HO from 5G to 4G:

```
rewards "to4G"
    [update] hoInter & in5G: 1;
endrewards
```

Here "hoInter" and "in5G" are *guards* which are boolean formulas that are true when a user makes inter-HO while it is currently present in 5G network. The value of the reward (one in this case) is given on the right side of the guard. This reward structure is executed whenever the action [update] is triggered in the model. Now the following LTL (linear temporal logic) formula is used to calculate the reward for the above mentioned structure:

$$\varphi_1 := R\{\text{"to4G"}\}max =?\ [C \leq T]$$

Here, C represents cumulative reward calculated for the time up to T. In addition to rewards, LTL formula can also be used to find the probability of an event. For example, the following formula captures the probability of current link type lt_c to be *outage* in the future.

$$\varphi_2 := Pmax =?[F\ \ lt_c = outage]$$

Here F is the temporal operator that represents any time instance in the future. Interested readers may refer to [22] for the detailed semantics of all the temporal operators used in LTL. In general, the output of the model checker is either reward-based HO calculation or the probability with which M satisfies φ. If system specifications are not satisfied by the model ($M \not\models \varphi$) within given constraints (such as probabilities, network/RF parameters, and discount factors), the input parameters are tuned to meet desired application objectives.

3) ZONE-BASED TRAFFIC SPLIT

Based on all the input parameters, the output of the model checker is HO count for varying values of velocity and gNB density. From these values of HO count and given a strategy for an application, traffic distribution is calculated and a UE is assigned appropriate zone that depends on its velocity, density of gNBs, and application requirement. The assigned zone determines how the traffic is split between MN and SNs. In particular, the model either suggests a user to make inter- (or intra-) RAT HO, or it triggers packet split between MN and SN if the number of HO count exceeds a threshold set by traffic distribution strategy.

The traffic split strategy is based on HO count which is directly related to the number of gNBs and the user velocity. Firstly, the model estimates HO count matrix with varying gNB count and user velocities. Secondly, the strategy is used to decide the trigger point for packet split mechanism. For instance, in aggressive strategy the traffic split starts at low values of HO count whereas in coward strategy the traffic split is triggered at high values of HO count. Once the split packet mechanism is initiated, specific percentage of traffic is routed towards MN.

For our DC-based model, we need activation function which helps to start, stop and manage share of data split mechanism between SN and MN. So, we select sigmoid function as activation of traffic split process as sigmoid exist between 0 and 1 values. The values between these two extremes are used to design linear traffic split zones between SN and MN. The sigmoid function is distributed into n (> 3) zones. Each zone is associated with traffic split discount factor, which defines the percentage through which data packets are routed towards MN. The traffic split discount factors δ_i for each zone z_i is defined as follows:

$$\delta_i(z_i) = \begin{cases} 0 & i = 1 \\ 1 & i = n \\ \Delta_h + \dfrac{(i-n+1)(\Delta_h - \Delta_l)}{n-3} & 2 \leq i \leq n-1 \end{cases} \tag{26}$$

Here Δ_h and Δ_l are the highest and lowest discount factors, respectively in traffic split scenario. Zone one (z_1) represents the scenario where HO count is less than the triggering point of split packet mechanism and hence no data packet is routed towards MN. The last zone (z_n) represents the other extreme, where HO count is very high and the algorithm routes all data packets towards MN. The process flow of these zone is shown in Fig. 5. A UE stays in LTE state as long as reward of 4G is higher than that of 5G. On the other hand, UE moves to 5G state where zones are calculated. When zone is greater than one, the UE moves to DC split state, otherwise it continues to stay in 5G state where all the data goes through SNs. When UE is in DC split state, it moves back to LTE or 5G state when zone changes to n or 1, respectively.

A script (written in MATLAB) is used that translates all zone formations into equivalent code in PRISM language. Zone-based traffic split strategy is used as input to PMC that verifies whether the split mechanism achieves the application objectives.

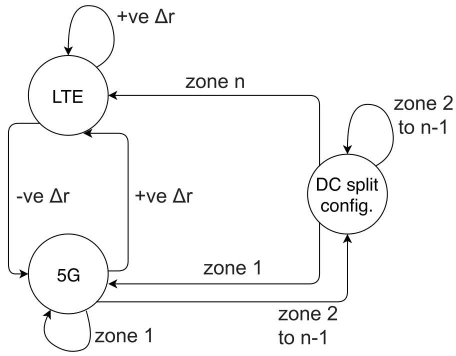


FIGURE 5. State transition diagram of the data split mechanism.

TABLE 3. Key parameters and their values used for evaluation.

Parameter	Values
Velocity (v)	15 - 35 m/s
Density of 5G BSs (ρ_{5g})	1 - 10
4G footprint (β_{4g})	1000 m
5G footprint (β_{5g})	100 - 300 m
Overhead due to intra-HO (α_a)	0.1
Overhead due to inter-HO (α_i)	0.2
4G BS bandwidth (BW_{4g})	20 MHz
5G BS bandwidth (BW_{5g})	200 MHz
Light load discount (δ_l)	0.5
Medium load discount (δ_m)	0.7
5G user area capacity factor (c_a)	1000
Traffic split discount factors (Δ_h, Δ_l)	0.85, 0.35
Total number of zones (n)	8

IV. RESULTS AND DISCUSSION

The proposed framework has been studied for varying density values of gNBs and user mobility profiles and the obtained results are presented in this section. The related parameters with their values are given in Table 3. We consider the situation where the coverage area is completely served by a 4G cell while the 5G coverage area varies and depends on the corresponding gNB profile, that is, density and footprint values. With the 4G cell coverage of 1000 m, we analyze the effect of the 5G network by increasing the number of gNBs with a coverage footprint of 200 m for each gNB.

For the aforementioned scenario, we consider the following two major cases: (1) multi-RAT MM without split configuration and (2) multi-RAT MM with split configuration.

A. MULTI-RAT MM WITHOUT SPLIT CONFIGURATION

In this case, we study the percentage of UEs connected to each RAT while increasing the 5G coverage. For this purpose,

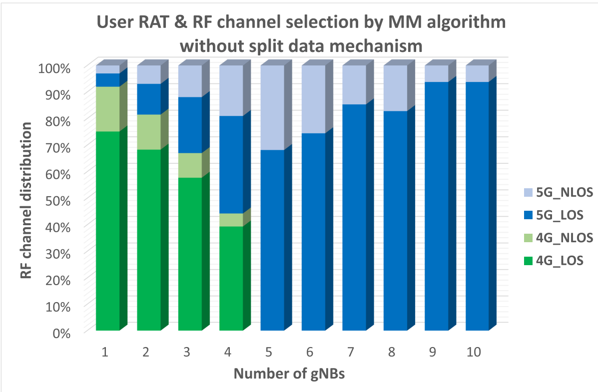


FIGURE 6. Impact of gNB densification on user RAT and associated RF channel state.

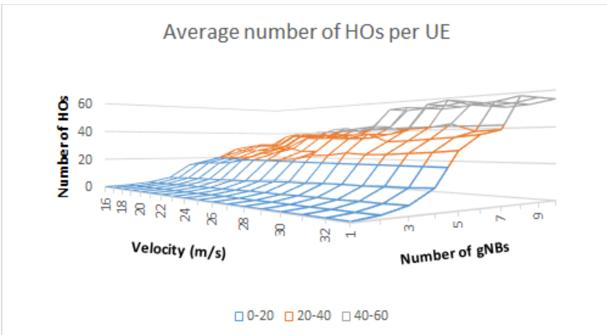


FIGURE 7. Impact of user velocity and gNB densification on intra 5G HOs.

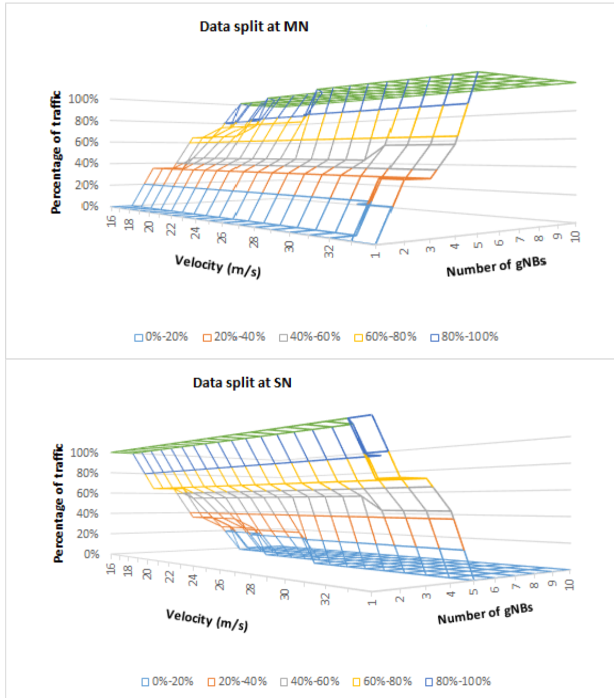


FIGURE 8. Data split based on aggressive strategy.

we increase the number of gNBs from one (sparse 5G coverage) to ten (dense 5G coverage). Fig. 6 shows the percentage of users attached to each channel type (*LOS*, *NLOS*) in 4G and 5G RATs for varying number of gNBs. Initially, with the presence of only one gNB in the scenario, most of the users are attached to 4G RAT. Since there is no blanket coverage for 5G RAT within the footprint of the 4G cell, 5G user penetration is limited due to coverage constraints.

With the increase in the number of gNBs, users get more 5G coverage and the proposed algorithm pushes more users to 5G RAT. When the number of gNBs reaches five, the available 5G coverage spans the entire footprint of 4G cell and the algorithm pushes all the users to 5G RAT. Increasing the number of gNBs beyond five simply results in an increased number of users attached to *LOS* channel within the same RAT. This can be seen in Fig. 6 that with the increased number of gNBs, the percentage of 5G *NLOS* channels that the users are attached to gets reduced.

The increase in the percentage of available *LOS* channels ensures good user experience and results in increased RF channel reliability. However, as the number of gNBs increases, the number of 5G intra-RAT HOs increases drastically for high mobility users as shown in Fig. 7. These frequent HOs cause bad user experience despite having a high-quality channel (*LOS*) in 5G RAT. The next subsection discusses the split configuration to alleviate this problem.

B. MULTI-RAT MM WITH SPLIT CONFIGURATION

In order to reduce the high number of HOs due to user velocity and gNB density, we employ aforementioned sigmoid-based methodology to split users' data between MN and SN. To demonstrate the functionality of the data split, we define two strategies, namely coward and aggressive. We first consider two extreme strategies in the following subsections. Later, we consider a set of strategies with another mobility scheme and compare system performance.

1) DATA SPLIT: AGGRESSIVE STRATEGY

In this strategy, data split toward MN is very aggressive and therefore the user will get most of the data through MN instead of SN. This strategy is very sensitive to total HO count and accordingly splits data towards MN very aggressively to avoid high number of HOs at SN layer. To demonstrate this strategy, we calculate the percentage of traffic split between 4G and 5G for varying values of gNB density (1 to 10) and UE velocity (16 to 34 m/s). We have set the threshold for this strategy to 10 HOs, after which the zone-based split mechanism distributes the packets to MN and SNs. The obtained results of data split between SN and MN for aggressive strategy are plotted in Fig. 8. We also provide 2D heatmap of zone formation for this strategy as shown in Fig. 9. It can be seen from the figures that both the UE velocity as well as the number of gNBs play an important role in deciding how much traffic will be routed through MN. Further, for users moving with relatively higher velocity, more traffic gets routed through MN even when there are fewer number of gNBs. In the extreme case when the UE velocity and the number of

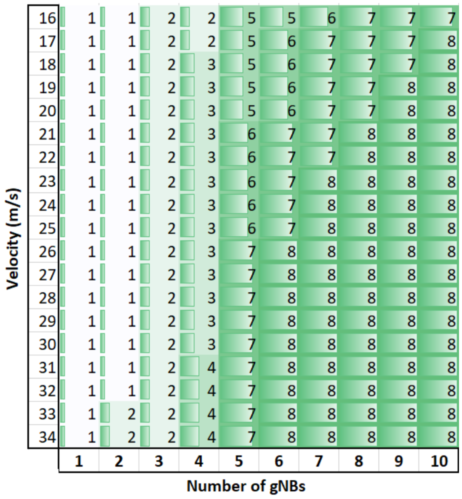


FIGURE 9. Heatmap showing zones in aggressive strategy.

gNBs both are high, there will be a huge number of HOs. This results in entire data transfer through MN with no traffic on the SN user plane. The impact of aggressive data split towards MN on system performance, particularly network throughput, will be covered in subsequent subsection.

2) DATA SPLIT: COWARD STRATEGY

In this strategy, data split towards MN is very small and hence users get most of the data through SN. We have set the threshold for this strategy to 40 HOs, after which the zone-based split mechanism distributes the packets to MN and SNs. Figure 10 shows data split for the coward strategy for varying values of velocity and gNBs, and the corresponding 2D heatmap of different zones is shown in Fig. 11. In this strategy, irrespective of the total HO count, there is no avenue where all SN traffic shifts to MN.

Similar DC-based MM scheme is proposed in a recent work [23] where high velocity users are treated as follows. First, activate the dual cell mode so that users can get connection from both macro and micro cells. Second, define threshold for high velocity users and accordingly move all high velocity users from 5G small cells to 4G macro cell. The corresponding data split towards MN and SNs are shown in Fig. 12. Rather than pushing certain percentage of users to 4G layer, our proposed split traffic mechanism routes some of its traffic towards MN. This results in improved user perception while keeping its traffic shared (based on sigmoid zones) at high capacity and high spectral efficiency 5G layer as discussed in the next subsection.

3) COMPARATIVE ANALYSIS

This subsection compares our methodology with one of the latest schemes proposed in [11] where the authors propose FHM (fast handover mitigation) scheme for high velocity users between 10 m/s to 40 m/s. The proposed algorithm reduces 79.56% of the total number of HOs in the network.

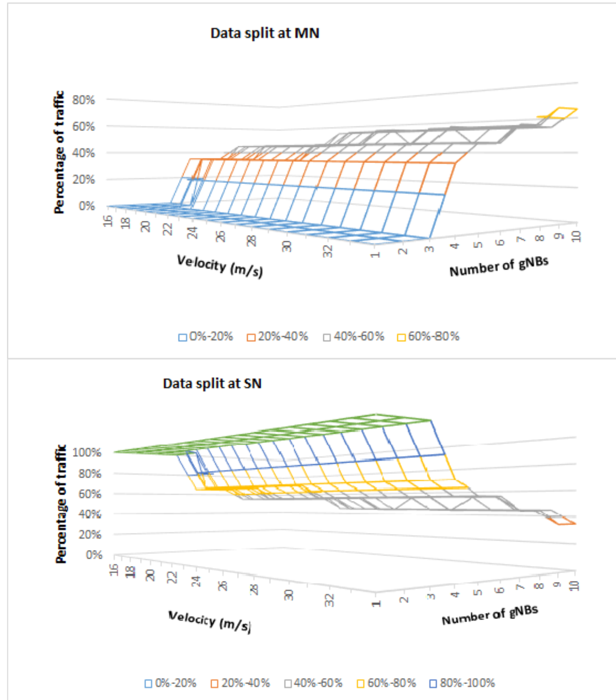


FIGURE 10. Data split based on coward strategy.

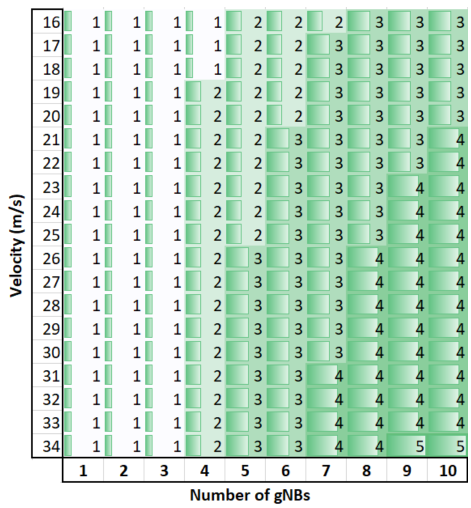


FIGURE 11. Heatmap showing zones in coward strategy.

Without FHM, the network suffers from many ping-pongs (due to unnecessary HOs and small sojourn time) and hence results in decreased throughput because unnecessary HOs consume small cell resources. With FHM, network resources are carefully utilized and hence network throughput increases by 10.82%. Although FHM achieves improved throughput, it suffers from the following issues:

1) The total number of users are assumed to have the following distribution: (a) 80% are normal users, b) 10% are ping-pong users, and (c) 10% are high velocity users.

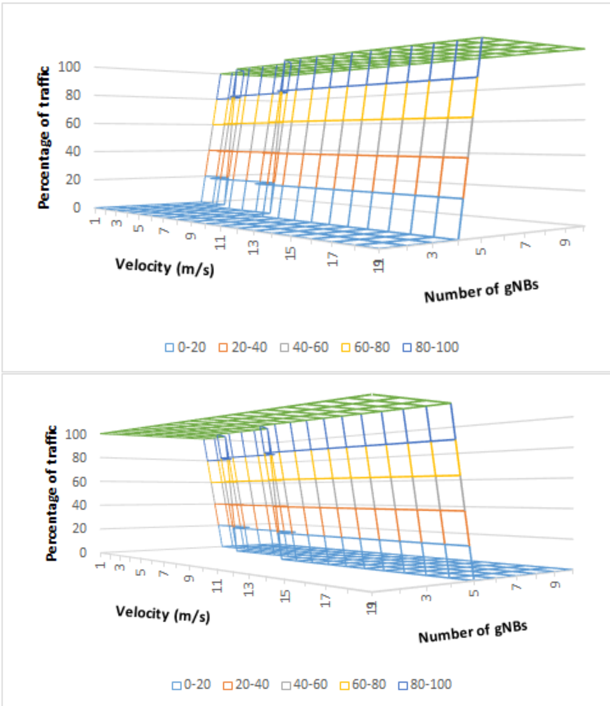


FIGURE 12. **Data split at MN (above) and SNs (below).**

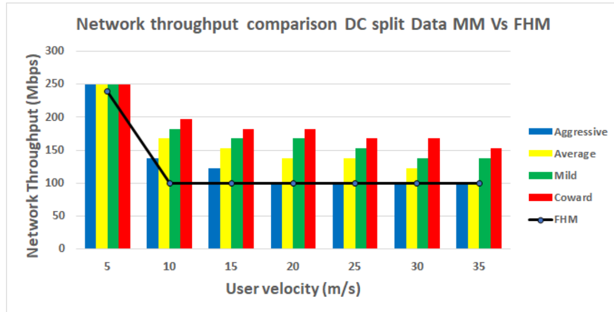


FIGURE 13. **Throughput for FHM and various strategies in DC-split configuration at varying velocity.**

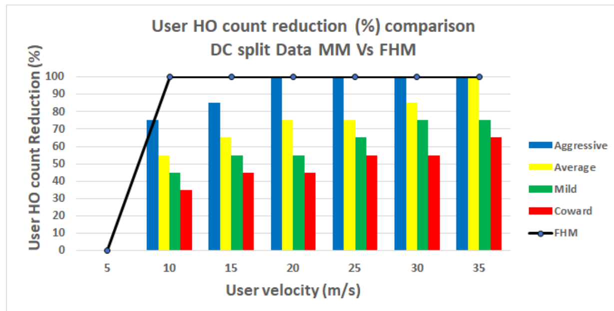


FIGURE 14. **Percentage of reduced HO count for FHM and various strategies in DC-split configuration at varying velocity.**

2) FHM achieves its goals by pushing 10% of high velocity users to 4G through hard HO.
3) Since FHM pushes 10% of high velocity users to 4G, these users suffer from sudden drop in throughput due to low capacity and less spectral efficiency of 4G layer. That is why, FHM scheme starts showing throughput degradation once the percentage of either total number of user or high velocity users increases.

To address the hard HO limitation of FHM, our proposed MM algorithm utilizes the strength of DC for MM as DC inherently benefits from reduced HO interruption as compared to interruptions due to conventional hard HO. To address sudden throughput degradation of FHM, our MM scheme employs appropriate data split mechanism between 4G and 5G RATs based on specific application strategy.

In this section, we present the comparative analysis of our DC-split data MM scheme with FHM. For this purpose, we use the same system parameters as that of FHM, that is, 50 users with 5 Mbps GBR (guaranteed bit rate) and 2 Mbps GBR data traffic per user at 5G and 4G layers, respectively. In FHM scheme, the network throughput value is 240 Mbps as 90% of the users (80% normal velocity users and 10% ping-pong users) continuously stay at 5G while only 10% high velocity users move to 4G layer. The 240 Mbps throughput starts to decrease when percentage share of high velocity users increases from 10% value. To capture this scenario, we take all 50 users under high speed category with 240 Mbps throughput when they stay at 5G and have velocity less than 10 m/s. Once their velocity reaches to certain threshold (selected as 10 m/s in this scenario), users move to 4G layer and hence experience sudden throughput decrease of 100 Mbps.

Our model splits certain percentage of the traffic towards 4G layer via MN while the rest of the traffic gets routed towards 5G layer via SN. This data split mechanism is shown in Fig. 13 for four different strategies *aggressive*, *average*, *mild*, and *coward*, with corresponding HO threshold values set to be 10, 20, 30, and 40, respectively. For aggressive strategy, more traffic is routed to MN which leads to low user throughput. Aggressive strategy behaves somewhat similar to FHM. On the other extreme, coward strategy avoids large data split towards MN while keeping more traffic at SN layer. This leads to better user throughput.

The reduction in HO count at 5G layer is another indicator for performance benchmarking between the proposed method and FHM. FHM shows HO count reduction up to 100% as users are pushed completely to 4G layer in this scheme. In the proposed scheme, the reduction in HO count depends on data split zone and strategy. The percentage of traffic that gets routed towards MN due to split mechanism does not experience high number of 5G layer HOs. As shown in the Fig. 14, reduction in the number of HOs is highest in aggressive strategy. This strategy performs almost similar to FHM in terms of HO count reduction. Aggressive strategy shows similar behavior in throughput comparison as well. Coward strategy, on the other hand, shows least percentage of reduced HO count as compared to other strategies as it routes smallest data share towards MN.

V. CONCLUSION AND FUTURE WORK

In this research work, we have developed a formal model of multi-RAT dual connectivity MM scheme for two-layer HetNet. The proposed MDP-based MM algorithm selects appropriate RAT during HO based on various parameters which include RF conditions, BS capacity, and the number of users on BS. The model also captures HO scenario for high velocity users by exploiting split data configuration mechanism. The framework incorporates advance features of 5G, like DC which has lower HO interruption as compared to convention hard HO. The split data configuration helps in routing certain percentage of traffic through alternate path in DC scenario.

We have evaluated the performance of the proposed framework with other MM schemes. The results show that the proposed model outperforms in terms of maintaining better network throughput. It also results in graceful reduction of total HO count at 5G layer. We have also presented the zone-based data split mechanism for different strategies. Moreover, the results show the flexibility in the presented framework which can be used to cater diverse QoS demands of 5G verticals.

In future, we plan to extend the framework to include other radio access layer processes such as RRM (radio resource management). The existing MM functionalities of the current framework with additional RRM will help capture a wide variety of 5G application scenarios.

REFERENCES

[1] M. D. Cia, F. Mason, D. Peron, F. Chiariotti, M. Polese, T. Mahmoodi, M. Zorzi, and A. Zanella, "Using smart city data in 5G self-organizing networks," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 645–654, Apr. 2018.

[2] A. Prasad, M. A. Uusitalo, D. Navrátil, and M. Säily, "Challenges for enabling virtual reality broadcast using 5G small cell network," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2018, pp. 220–225.

[3] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5G: RAN, core network and caching solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3098–3130, 4th Quart., 2018.

[4] A. Jain, E. Lopez-Aguilera, and I. Demirkol, "Are mobility management solutions ready for 5G?" 2019, *arXiv:1902.02679*. [Online]. Available: http://arxiv.org/abs/1902.02679

[5] Y. Li, B. Cao, and C. Wang, "Handover schemes in heterogeneous LTE networks: Challenges and opportunities," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 112–117, Apr. 2016.

[6] A. S. Cacciapuoti, "Mobility-aware user association for 5G mmWave networks," *IEEE Access*, vol. 5, pp. 21497–21507, 2017.

[7] R. Arshad, H. Elsawy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Handover management in 5G and beyond: A topology aware skipping approach," *IEEE Access*, vol. 4, pp. 9073–9081, 2016.

[8] Z. Wang, L. Li, Y. Xu, H. Tian, and S. Cui, "Handover control in wireless systems via asynchronous multiuser deep reinforcement learning," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4296–4307, Dec. 2018.

[9] A. Gharsallah, F. Zarai, and M. Neji, "SDN/NFV-based handover management approach for ultradense 5G mobile networks," *Int. J. Commun. Syst.*, vol. 32, no. 17, p. e3831, Nov. 2019.

[10] T. Mumtaz, S. Muhammad, N. Mohammad, M. I. Aslam, and I. Ahmed, "Modeling and evaluation of mobility management in mmWave cellular networks," in *Proc. IEEE 21st Int. Multi-Topic Conf. (INMIC)*, Nov. 2018, pp. 1–6.

[11] M. M. Hasan, S. Kwon, and S. Oh, "Frequent-handover mitigation in ultra-dense heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1035–1040, Jan. 2019.

[12] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmWave mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2069–2084, Sep. 2017.

[13] R. P. Antonioli, E. B. Rodrigues, D. A. Sousa, I. M. Guerreiro, C. F. M. e Silva, and F. R. P. Cavalcanti, "Adaptive bearer split control for 5G multi-RAT scenarios with dual connectivity," *Comput. Netw.*, vol. 161, pp. 183–196, Oct. 2019.

[14] *Study on New Radio (NR) Access Technology*, document TS 38.912, Version 15.0.0, 3GPP, Jul. 2018. [Online]. Available: https://www:3gpp:org/ftp/Specs/html-info/38912:htm

[15] *NR; Multi-Connectivity; Overall Description; Stage-2*, document TS 37.340, Version 16.0.0, 3GPP, Jan. 2020. [Online]. Available: https://www.3gpp.org/ftp/Specs/html-info/37340.htm

[16] *NR; User Equipment (UE) Radio Transmission and Reception*, document TS 38.101, 3GPP, Jan. 2020. [Online]. Available: https://www.3gpp.org/ftp/Specs/html-info/38101.htm

[17] C. Baier and J.-P. Katoen, *Principles of Model Checking* (Representation and Mind Series). Cambridge, MA, USA: MIT Press, 2008.

[18] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks-with a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, 2017.

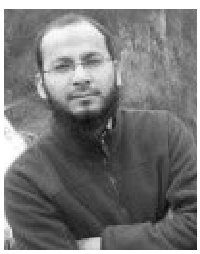
[19] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document TS 38.901, Version 16.0.0, 3GPP, Jan. 2020. [Online]. Available: https://www.3gpp.org/ftp/Specs/html-info/38901.htm

[20] S. Jaeckel, M. Peter, K. Sakaguchi, W. Keusgen, and J. Medbo, "5G channel models in mm-wave frequency bands," in *Proc. 22th Eur. Wireless Conf. Eur. Wireless*, 2016, pp. 1–6.

[21] *5G Study on Scenarios and Requirements for Next Generation Access Technologies, Release 15*, document TS 38.913, Version 15.0.0, 3GPP, Jul. 2018. [Online]. Available: https://www.3gpp.org/ftp/Specs/html-info/38913.htm

[22] M. Kwiatkowska, G. Norman, and D. Parker, "PRISM 4.0: Verification of probabilistic real-time systems," in *Proc. 23rd Int. Conf. Comput. Aided Verification (CAV)*, in Lecture Notes in Computer Science, vol. 6806, G. Gopalakrishnan and S. Qadeer, Eds. Berlin, Germany: Springer, 2011, pp. 585–591.

[23] M. Tayyab, X. Gelabert, and R. Jäntti, "A survey on handover management: From LTE to NR," *IEEE Access*, vol. 7, pp. 118907–118930, 2019.



TARIQ MUMTAZ received the B.E. degree in electrical engineering from the NED University of Engineering and Technology, Pakistan, in 1999, and the M.S. degree from Arizona State University, USA, in 2002. He is currently pursuing the Ph.D. degree with the NED University of Engineering and Technology. He worked as a Mobile Network Consultant of different telecom markets in Asia, the Middle East, and North America. He has extensive RF hands-on experience of different vendor's equipment such as Ericsson, NSN, Huawei, and ALU. He is currently working as a Lecturer with the School of Science and Technology, Habib University, Pakistan. His areas of research include radio resource management and cross-layer optimization for 5G market verticals.

SHAHABUDDIN MUHAMMAD received the B.E. degree from the NED University of Engineering and Technology, Pakistan, in 1999, and the Ph.D. degree from the University of Central Florida, Orlando, USA, in 2007. He has extensive software development experience at various companies in the USA, including Adaptec, Orlando, FL, USA, Riverbed Technologies, San Francisco, CA, USA, and Microsoft, Redmond, WA, USA. He is currently working as an Assistant Professor with the College of Computer Engineering and Science, Prince Mohammad Bin Fahd University, Saudi Arabia. His research interests include formal verification, sensor networks, security protocols, mobile communication, and the Internet of Things.

MUHAMMAD IMRAN ASLAM was born in Karachi, Pakistan. He received the B.E. and M.Eng. degrees in electrical engineering from the NED University of Engineering and Technology, Karachi, in 2001 and 2005, respectively, and the Ph.D. degree in electrical engineering from Michigan Technological University, Houghton, MI, USA, in 2012. He is currently serving as an Associate Professor with the NED University of Engineering and Technology. His research interests include wireless communication, electromagnetics, optical metamaterials, and green telecommunication.

NAZEERUDDIN MOHAMMAD received the B.E. degree in electronics and communications engineering from Osmania University, India, in 1996, the M.S. degree in systems engineering from the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 1999, and the Ph.D. degree from the School of Computing and Information Engineering, University of Ulster, Coleraine, U.K., in 2007. He is currently working as the Director of the Cybersecurity Center, Prince Mohammad Bin Fahd University, Saudi Arabia. Prior to the current position, he worked as an IT Infrastructure Manager/Administrator and a Lecturer at the KFUPM, a Postgraduate Researcher at the University of Ulster, a Research Intern at Cisco Systems, USA, and an IT Infrastructure Manager/Administrator and a Researcher at the Portland House Research Group, Australia. He was a recipient of the M.S. degree Research Scholarship from the KFUPM, in 1997, the Vice Chancellor's Research Scholarship from the University of Ulster, in 2004, and the First Prize of the Faculty of Engineering Business Plan Competition from the University of Ulster, in 2006. He is actively involved in the TPC/review committees of renowned conferences.