This article has been accepted for inclusion in a future issue of this magazine.

# A Framework to Address Mobility Management Challenges in Emerging Networks

Asad Zaidi[1], Hasan Farooq[2], Ali Rizwan[3], Adnan Abu-Dayya[4], Ali Imran[2]

Electrical and Computer Engineering, The University of Oklahoma, Tulsa, USA[1]

Ericsson Research, Santa Clara, CA, USA[2]

Qatar Mobility Innovations Center[3]

Department of Electrical Engineering, Qatar University, Doha Qatar[4]

Email: {asad, ali.imran} @ou.edu[1]

hasan.farooq@ericsson.com[2]

arizwan@qmic.com[3]

adnan@qu.edu[4]

*Abstract*- The key enablers for emerging cellular networks such as densification, concurrent operation at multiple bands and harnessing mmWave spectrum give birth to a peculiar set of new network management challenges. One such key challenge is the user mobility management. In this article, we identify the key issues that render current mobility management paradigm inadequate for delivering the expected Quality of Experience (QoE) and resource efficiency in emerging and future cellular networks. Together, these challenges call for paradigm shift in the way mobility is managed in cellular networks. We present an Advanced Mobility Management and Utilization Framework (A-MMUF) that can enable this paradigm shift by transforming mobility management from being a reactive to a proactive process. The core idea of A-MMUF is to build upon Mobility Prediction Models (MPMs) to predict various attributes of user mobility and traffic patterns such as next candidate cell for handover (HO), time of HO, future cell loads. These predictions are then leveraged to not only improve HO process for better QoE and less signaling overhead, but also to enable proactive automation to further maximize network performance in terms of energy and spectrum efficiency. However, understandably, the gains offered by the A-MMUF hinge on the accuracy of the MPMs, which may vary, not only with the choice of underlying machine learning technique, but also with the volume, variety, and fidelity of the training data. We analyze the potential gains of A-MMUF through three case studies: namely proactive HO, proactive Mobility Load Balancing (P-MLB), and proactive Energy Savings (P-ES). In addition to demonstrating significant gains in all three case studies, the results provide useful insights into the agility vs accuracy tradeoff that can be leveraged for choosing optimal machine learning models for practical deployment of A-MMUF.

## I. Introduction

Mobility is *raison d'etre* of wireless cellular networks. The ability to seamlessly support User Equipment (UE) mobility has been the key factor behind the wide scale penetration and applications of the wireless cellular networks. A recent study projects that mobile data traffic will increase by 287 times between 2017 and 2027, where 5G will account for 90% of mobile subscriptions in North America [1]. This trend combined with imminent advent of autonomous cars and unmanned aerial vehicles highlights the need for making mobility management in future networks even more efficient, reliable, and seamless. On the other hand, network densification coupled with the millimeter-wave (mmWave) bands adaptation is emerging as the predominant mechanism to meet the capacity enhancement goals [2]. Emerging networks are envisioned to be based on heterogeneous (HetNets) architectures where the ubiquitous and continuous connectivity will be rendered by the legacy/advanced sub-6GHz macro cells and the additional capacity will be provided by the deployment of a large number of small-cells, operating at higher bands. Particularly, mmWave small-cells are being considered as the game-changer for the future ultra-dense multi-band networks. Deploying more mmWave Base Stations (BSs) within the same geographical region has strong potential to solve the two long-standing and intertwined problems in cellular networks of spectrum scarcity and interference.

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.

While the ultra-dense deployment along with mmWave has potential to address the capacity crunch, it leads to a set of network management problems. One such key yet largely overlooked problem is the compounded increase in the number of handovers (HO) [3]. While the control-plane is provided by sub-6GHz cells, the transition of user-plane from one small-cell to another small-cell in a dense network deployment scenario is an open problem. Similarly, for semi-static users, load balancing from one small-cell to another small-cell is another big challenge. This problem alone leads to numerous network resource efficiency as well as user Quality-of-Experience (QoE) issues such as: increased signaling overhead and hence poor spectrum utilization, higher latency, decreased average UE throughput and battery life due to time lost in cell discovery, increased chance of radio-link and HO failures. In fact, 3GPP studies [4] reveal that HO failure rate in the traditional HetNets is as high as 60%. The average failure rate almost doubles as compared to one in a macro-only network. A detail description of intra- and inter-frequency HO along with the associated challenges is given in [3] [5].

In the next section we explicate the challenges in mobility management in Ultra-Dense Multi-Band Network (UDMN) that must be addressed to meet resource efficiency and QoE in future networks. In the subsequent sections, we present an advanced mobility management framework called Advanced Mobility Management and Utilization Framework (A-MMUF) that has potential to address the aforementioned challenges. The key novelty of A-MMUF lies in building UE Mobility Prediction Models (MPMs) and then exploiting them to not only improve HO process in UDMN, but also to enable proactive self-optimization. Through three case studies on proactive HO, proactive Mobility-Load-Balancing (MLB) and proactive Energy-Savings (ES), we show how A-MMUF can help achieve proactive mobility management and enable numerous proactive Self-Organizing-Network (SON) functions, now also called zero touch automation and will be referred as automation in rest of the paper. As the gains offered by A-MMUF hinge on the accuracy of the MPMs and in real networks no MPM may yield 100% accuracy, we also analyze the impact of MPMs accuracy on potential gains achievable by A-MMUF under various scenarios.

## II. Challenges in mobility management of UDMN

The following challenges define the breadth and depth of the impending yet under-addressed problem of mobility management:

### *1. Resource-Inefficiency and Scarcity*

#### 1.1. Resource-Hungry Mobility Support

In current cellular networks that mainly consist of macro cells on sub-6GHz bands, mobility support is offered by incorporating elaborate HO mechanisms that rely on heavy pre- and post-HO signaling (preparation, execution, and completion phases). Thus, supporting UE mobility is already a resource-hungry and extremely complex process in legacy networks that causes significant signaling overhead. With extreme densification, number of HOs increase dramatically and mobility management signaling overheads start offsetting the capacity gains.

#### 1.2. Large HO Completion Time

The time to complete a HO is targeted at 65ms [6] in a 4G network, and this time is designed mainly for macro cell-based deployments. However, in UDMN, given the much smaller average cell size and thus the small UE sojourn time, the time to complete a HO must be reduced significantly. The new agile HO design also needed to meet the ambitious low latency requirements imposed by 5G/6G use-case of Ultra Reliable Low Latency Communications (URLLC) [7].

#### 1.3. Mobile Battery life

To perform a HO in UDMN, mobile devices must discover small-cells operating on different frequency bands by periodically running an Inter-frequency small-cell discovery process. UDMN will require inter-frequency small-cell discovery rate to be much higher than the current rate for 4G. This will exacerbate mobile battery life problem in UDMN.

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore.  Restrictions apply.

*2. User-Experience Degradation*

2.1. High risk of Breakdown of Current Mobility Management Mechanism

HO failure can be caused by many factors such as loss of signaling during HO or during admission control where the target cell does not have resources for incoming bearers. Because of HO failure, **Radio Resource Control connection** needs to be re-established which leads to extra control signaling overhead and degraded QoE. In 4G networks, HO failure rate is targeted at less than 5%. However, 3GPP study [4] shows that adding only ten small-cells per macro cell can push the HO failure rate as high as 60%, indicating the potential breakdown of current mobility management mechanism in UDMN.

2.2. QoE Degradation During Intra-Frequency Handover

Another caveat of increased HO rate is the hidden loss in average throughput and higher chances of radio-link-failure. This is primarily due to the temporary negative Signal to Interference and Noise Ratio (SINR in dB scale) just before the HO trigger, where the HO target cell is stronger than the serving cell by more than or equal to HO margin, and for a duration equal to sum of a) timeToTrigger [6] [5] and b) the delay due to Random Access Channel and HO processing at target BS. The resultant SINR dilapidation becomes more critical with the speed of the UE) or due to the suboptimal HO parameter configuration or cell overlap planning. The degraded SINR condition reduces the UE throughput and results in an increase in data retransmission rate. In addition, the higher chances of RLF led to increased HO failures and negatively affects retainability Key Performance Indicator (KPI).

2.3. Loss of Throughput During Inter-Frequency Handover

Given planned exploitation of a diverse range of bands, inter-frequency mobility is a vital component of 5G and 6G but has not received the attention it deserves in the research community. Inter-frequency mobility requires event A2 to be triggered, which is followed by the configuration of periodical measurement gap for the UE by the BS. However, this process interrupts data transmission and reception as the UE shifts the radio to measure reference signal strength from the neighboring cells operating at a different frequency. As a result, UE cannot transmit/receive data for 10ms resulting in up-to 25% loss in throughput. UDMNs with a variety of frequencies ranging from sub-6GHz to mmWave band may require the UE to perform an extensive search of available frequencies before initiating a mobility decision. This issue can be aggravated when considering the latency goal of <1ms [7] for URLLC use-case. A similar issue can arise during E-UTRAN New-Radio Dual-Connectivity where UEs will traditionally camp on 4G and after searching for an optimal 5G cell, dual-connectivity activation process will be initiated.

*3. Unprecedented Channel Behavior in mmWave Cells*

3.1. Non-Graceful Signal Decay

Conventional cellular bands exhibit graceful signal decay. This graceful signal decay along with planned cell overlaps allows the use of HO Margin (*offset+hysteresis*) for HO preparation and execution phase and to avoid ping pong as shown in Fig. 1. However, mmWave cells in UDMN will have a sudden signal decay for example when link becomes non-line of sight. This means no or little cell overlaps leaving far less time to prepare and execute HOs. This challenge alone renders current mobility management paradigm inadequate for future networks and requires rethinking of the way HOs are performed.

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this magazine.

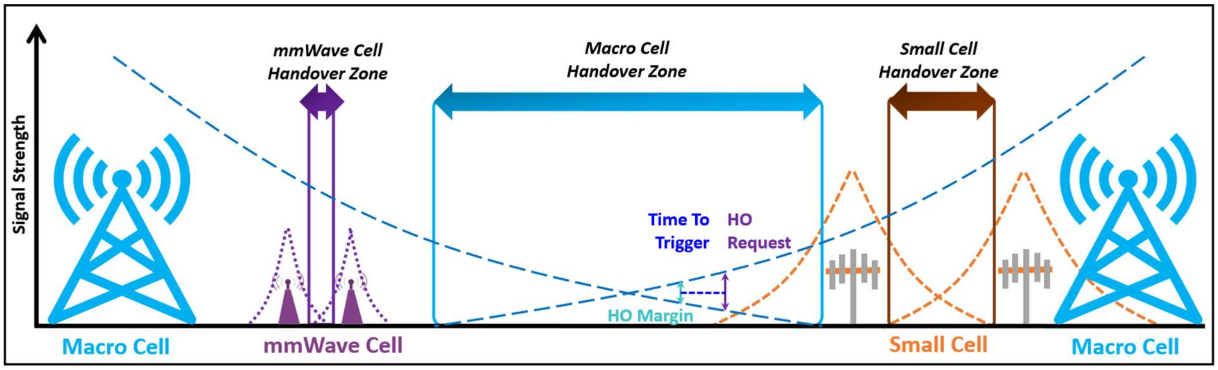


*Fig. 1. Relative spatial margins to complete HOs in UDMN consisting of macro cells, conventional small-cells and mmWave small-cells.*

3.2. Cell Discovery Issue with mmWave Small-Cells

Conventional sub-6 GHz band cells have omni-directional or wide-beam sector antennas, and thus can be easily discovered by an approaching mobile user to start the HO process. On the contrary, mmWave cells rely on narrow beams to overcome the high propagation losses and require line of sight conditions. This means unless a mmWave cell has aligned its beam with an oncoming mobile user using advanced beam control and beam management technologies, it cannot discover the user, or be discovered by the user to start the HO processes. This gives rise to an unprecedented cell/user discovery problem in mmWave cells making mobility management more challenging in UDMN as described in [2].

*4. Reactive Automation Use-Cases Affected by the Mobility Management in UDMN*

Finally, aforementioned idiosyncrasies of UDMN render ineffective a number of recently standardized automation functions such as the Mobility-Robustness-Optimization (MRO), MLB, and ES [8] . This is because the current design of these automation functions relies on assumption of relatively low rate of UE transitions among cells and thus relatively slow variation in cell loads that is reflective of macro cell only network.

Remarkably, until now most of the research towards UDMN remains focused on channel modelling and hardware design aspects of the mmWave based UDMN. The mobility management in UDMN so far remains a Terra incognita. Above challenges imply that, if no measures are taken to rethink mobility management for future UDMN, UE mobility management can become the bottleneck in QoE as well resource efficiency in practical deployments of UDMN despite of advances in physical layer, and adaptation of new spectrum.

In the subsequent sections, we present an Advanced Mobility Management and Utilization Framework (A-MMUF) for addressing the aforementioned challenges.

## III. A FRAMEWORK FOR ADVANCED MOBILITY MANAGEMENT AND UTILIZATION

The core idea of the advanced mobility management and utilization framework (A-MMUF) is illustrated in Fig. 2. The two cornerstone features that make A-MMUF distinct from existing mobility management paradigm are:

- Leveraging Artificial Intelligence (AI) to predict mobility patterns from data readily available in the network.
- Leveraging mobility prediction to enable proactive mobility management and proactive automation functions (P-AUTO).

The proposed A-MMUF framework highlighted in Fig. 2 involves the following steps:

1) _Training data collection:_ The sources of this data for training the mobility prediction models could be call detail records, Minimization of Drive Test measurements, HO reports, crowd source data, geographical maps etc. [8]. However, almost all real data is expected to be sparse or incomplete. A domain knowledge-based pre-processing and cleansing of the data is required before it can be used for model training. For example, from the domain knowledge (field experience and theory) we know that apart from shadowing, the amount of HO margin and the UE direction and UE velocity will offset the HO completion location compared to the

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this magazine.

actual cell border. The collected data therefore should have a buffer zone to accommodate such variations. Similarly, since the UE capability affects the user mobility direction when performing inter-frequency HO, the collected data should not be biased towards a particular set of frequencies. Moreover, since the radio conditions and corresponding radio link failures are highly dependent on LoS/NLoS and outdoor/indoor conditions, having a marker for these categories helps improve the quality of collected data. In case the data is small but representative, methods such Generative Adversarial Networks (GANs) can be used to augment the data. In cases where data is incomplete, it can be augmented using synthetic data such as that produced from well calibrated system level simulators with realistic mobility modeling [5].

2) *Designing and training the mobility prediction models:* The Mobility Prediction Model (MPM) used to predict a set of attributes of UE mobility should meet two criteria: 1) The MPMs should not only be accurate but also computationally low cost and they should be capable of performing on real time; 2) The MPMs act as enabler for the proactive automation functions in UDMN. A simple yet effective way to build such MPM is to leverage UE reported measurements, which include traces of past cell transitions such as cell IDs, HO failure reports, Reference-Signal-Received-Power measurement report and call-records to build and train low complexity MPM. This high frequency and large volume data need to be processed with advanced storage techniques as highlighted in Fig. 2. Discrete Time Markov Chains (DTMC) have been commonly leveraged in literature for MPM as they can yield more scalable solution because of being memory-less. DTMC does not need to store users' past movements, instead the crux of this information is captured by transition probabilities. Probability or likelihood of a UE to transit to next cell can be predicted by modelling UE transition from one cell to another as a Markov stochastic process. Each state in the DTMC represents a cell wherein HO from a cell to another can be modelled as state transition. However, the DTMC is limited to spatial prediction only. It does not take into account time domain. A solution to solve this problem is to use semi-Markov chain for building spatio-temporal based MPM models. A similar model for mobility prediction is presented in [9] which enables MRO, MLB and ES automation functions. A prediction accuracy of more than 80% has been achieved using semi-Markov model in experimental evaluation [10]. Moreover, the state of the art supervised machine learning techniques like Deep Neural Networks, Support Vector Machine, Decision Trees (XGBoost) can also be leveraged for mobility prediction to achieve better accuracy, albeit at the cost of increased training and prediction time as shown in recent study [10] that provides performance comparison of various AI tools for building a MPM from cellular data. To exploit various AI tools, user's trajectory can be represented as a set of cells it traverses to transform the mobility prediction problem into a classification problem with target cells as class labels [10]. Out of the available ML approaches for MPM models in literature [9] [11], selection for MPM to be used in practical deployment of A-MMUF in a given scenario can be made based on the desired level of tradeoff between several criteria that include accuracy, training time, prediction time, and quantity and quality of data required to train the model.

3) *Triggering HO and automation functions proactively:* The MPMs can then be used to predict various mobility metrics such as next HO cell, time of HO, direction of arrival, frequency of HO, load prediction etc. These predictions can be used to not only trigger HO proactively and mmWave cell discovery, but they can also make many automation functions proactive rather than reactive such as MRO, MLB, ES, Coverage and Capacity Optimization (CCO), Inter-Cell-Interference-Coordination (ICIC). Here, the former refers to Proactive-Mobility Management (P-MM) and the latter Proactive automation (previously known as Proactive SON), as shown in Fig. 2. For an instance, MPM can predict number of UEs to be associated with a cell in a future time interval. Thus, it can act as an enabler for transforming several reactive automation functions to be proactive e.g., energy saving algorithms can optimize the activation and de-activation of the cells based on the predicted load and thus enhance energy efficiency of the overall network. Similarly, if the predicted number of UEs to be associated with a cell in the next time interval is greater than a certain threshold the MPM can be the basis to proactively trigger MLB. Based on the information of estimated load, MLB can change HO parameters or coverage area of the to-be over-loaded cell and its neighbors such that the extra traffic is offloaded to neighboring cells and the load is evenly balanced among the cells without sacrificing QoE. Thus, MPM can transform the most of automation use-cases from reactive to proactive ones.

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this magazine.

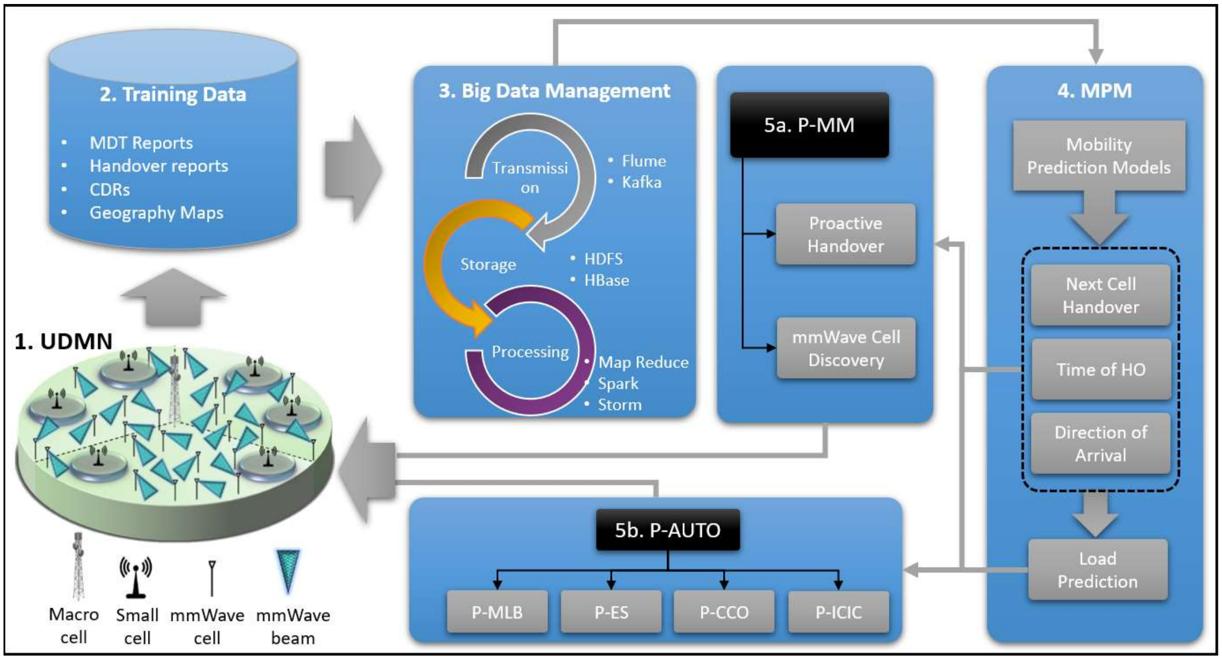


*Fig. 2. Our proposed Advanced Mobility Management and Utilization Framework (A-MMUF)*

However, it can be anticipated that the gains provided by A-MMUF will heavily depend on the accuracy of the MPMs. In the following section we present three case studies that analyze the gains of A-MMUF for varying accuracy of the MPM.

## IV. A-MMUF Case studies

In this section, we present three case studies of the presented framework (A-MMUF), namely proactive HO, proactive MLB (OPERA), and proactive-ES (AURORA).

### *A. Proactive HO Scheme: Minimizing HO Signaling Overhead and Latency in UDMN:*

In current networks, mobility is managed through an elaborate HO process that includes preparation, execution, and completion phase. Each of these three phases requires the exchange of numerous measurements [3] [11] from UE to the serving cell, and to the core network, causing huge signaling and latency. In UDMN, the problem will get exacerbated by the fact that unlike macro cells, small-cells are not typically directly connected to the core network with dedicated high-speed signaling interfaces where mobility procedures are usually coordinated. This, combined with the increased HO rate anticipated in UDMN and low latency requirement expected from 5G, call for transformative change in the way HOs are managed.

To address this challenge, MPMs can be an enabler for a proactive HO (P-HO) scheme. P-HO can be achieved by predicting the next cell and HO-time through MPM. Consequently, HO preparation phase can be done (e.g., advance resource allocation) before the HO criteria is satisfied reducing the HO induced latency. Moreover, HO execution phase can be made seamless by reserving access related resources in the target cell ahead of time. Finally, the HO completion part where the path switch of user plane from source to target cell takes place can be totally avoided by core network sending user plane directly to the target cell.

<u>Accuracy vs Agility Tradeoff and its impact on Proactive Automation</u>: The core idea of P-HO is not new, and its feasibility has already been studied in literature [3] [11]. Here we analyze the dependance of P-HO's gain (relative to conventional HO) on the attributes of underlying MPM to gain insights into the accuracy vs agility tradeoff. Simulation setup used for P-HO evaluation includes 130m inter-site distance and user density of 5UEs/small-cell with UEs moving at speed of 3km/h, 30km/h, and 100km/h. During simulation, UEs are configured with 2dB HO hysteresis and 200ms measurement gap [11]. From domain knowledge, we know that the UE penetrate the coverage of HO target cell during the HO criteria evaluation & this creates a HO buffer zone that varies with UE trajectory and UE velocity. We take this under consideration during the simulation and evaluation phase. Fig. 3 shows the gain achieved by P-HO by using several MPM predictors. We quantify the HO signaling latency (expressed in terms of the delay required

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this magazine.

to transmit and process the HO messages) reduction achieved through various MPMs following the approach in [11]. As it can be seen in Fig. 3, the P-HO gain increases with increase in accuracy of the underlying MPM reaching up to 22.38% using XGBoost based MPM. While in literature often only accuracy is emphasized, when important factors for achieving the agility needed for P-HO implementation in real network are considered such as training time and inference time, XGBoost based MPM is not the clear winner. In that case, Semi-Makrov model based MPM emerges as better choice as it can be retrained quickest, albeit at the cost of lower accuracy and hence lower P-HO gain. The results bring forth the accuracy vs agility tradeoff that often remains hidden in literature but has valuable insights for choosing an optimal MPM model for given deployment scenario. A given deployment scenario is characterized by availability of computational resources and non-stationarity of the UE mobility patterns. Non-stationary UE mobility patterns may require proportionally higher rate of retraining of the MPM, thus making less accurate, but more agile MPMs such as Semi-Markov based MPM can be a better choice compared to the most accurate models.

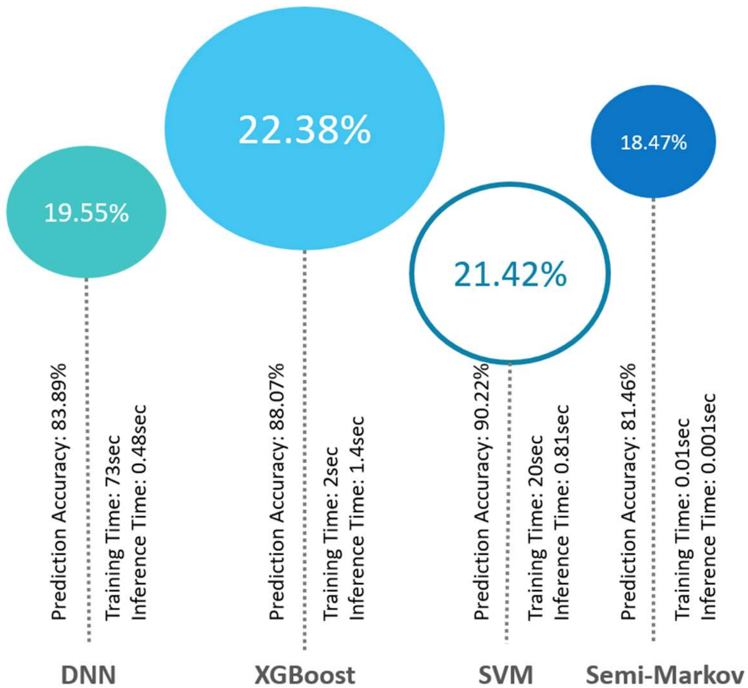


*Fig. 3. HO gain achieved by P-HO scheme based on various MPMs, in comparison to non-predictive HO scheme*

## *B. Proactive MLB*

MLB targets to optimally distribute traffic over the cells to avoid congestion and to maximize the resource efficiency in the face of acute UE mobility. However, with shrinking cell sizes and increased rate of UE transitions among cells, the cell loads in UDMN will change much faster than the rate of cell load change in macro cells. This means congestions and resource inefficiency both will be observed more often in UDMN if no counter measures are taken. Legacy load balancing schemes that get triggered from the congestion indicators, i.e., after the congestion has already happened, while adequate for current networks, will not suffice to meet QoE requirement in future networks. One way to address this problem is to transform MLB into Proactive-MLB (P-MLB). Authors in [12] propose OPERA framework where, if the predicted load of the cell in the next time interval is greater than certain threshold, P-MLB can adopt the HO, cell association and admission control parameters of that cell and its neighboring cells such that load in the cell to be congested will always remain pre-set threshold and congestion will not be observed at all. In [12], proactive congestion threat is detected using Semi Markov model. Congestion avoidance takes place by optimizing soft parameter (CIO) or hard parameters (azimuth, tilt, transmission-power, beamwidth).

Fig. 4 illustrates the benefits of P-MLB in comparison to reactive MLB approaches (RDS-A, RDS-B, RDS-C, and Joint). RDS-A, RDS-B, and RDS-C are the three most common configurations adopted from real network 4G deployment settings for one of USA's national mobile operators in the city of Tulsa. Similarly, in Joint algorithm, LB

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this magazine.

is achieved by changing tilts while keeping in view the coverage constraints. We use Monte Carlo style simulations for the evaluation of OPERA framework. Simulation setup includes 336 stationary UEs and 84 mobile UEs following the realistic SLAW mobility pattern, and the HO statistics is collected for a period of one week. Fig. 4 highlights the reduction in number of unsatisfied UEs and the gain in UE SINR achieved by employing P-MLB based OPERA framework. P-MLB can be a strong candidate which can help to realize the key 5G/6G use-case of URLLC through next cell prediction, timely resource allocation and ensuring good SINR.

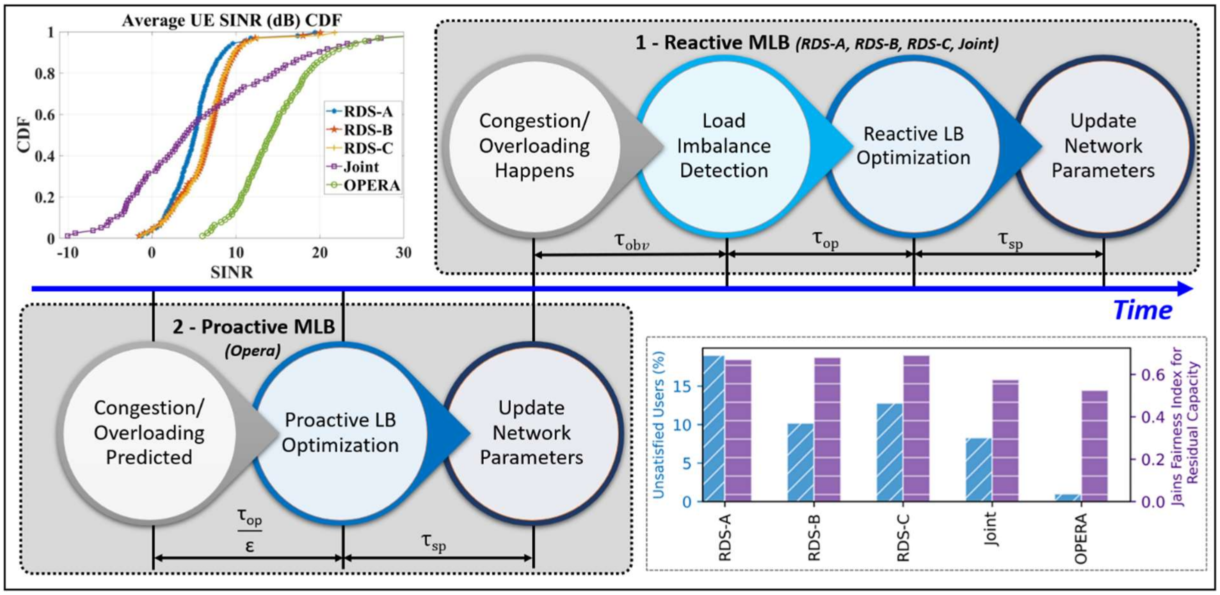


*Fig. 4. Gain in QoE achieved by using P-MLB based OPERA framework [12].*

### *C. Proactive Energy Savings (ES)*

The objective of ES automation use-case is to minimize the overall energy consumption for a given network load. Energy consumption of a mobile network can be obtained using the Energy-Consumption-Ratio (ECR) metric. ECR for a cell is defined as the amount of energy consumed in Joules per each bit of information that is reliably transmitted in that cell. During low traffic times, cells can be switched to sleep state with minimum energy consumption and the small traffic can be offloaded to the neighboring overlapping cells. These can be switched back ON during high traffic times. However, all current ES schemes suffer from two common problem: 1) switching a dormant cell back on has a latency that can lead to poor QoE; 2) How to discover an OFF cell. Both problems can be solved by leveraging MPM to transform ES into proactive ES scheme as in [13]. The key idea here is to predict future cell loads using past HO traces, and accordingly schedule the sleep cycles for the small-cells. This not only allow the sleeping cells to be turned ON proactively than reactively, but also minimizes the switching ON latency and addresses the cell discovery challenge.

The average Energy-Reduction-Gain defined in detail in [13] (ratio of difference between $ECR^{P\text{-}ES}$ scheme and $ECR^{allON}$, versus the $ECR^{allON}$) of proactive energy scheme for these mobility prediction schemes against conventional practice of keeping cells ON all the times is plotted in Fig. 5. Fig. 5 shows that the gain of proactive energy scheme increases with the prediction accuracy and is found to be maximum (80.68%) for XGBoost. The results also show like other P-AUTO use-cases discussed above, the ERG also heavily depends on the accuracy of the MPM. Simulation was conducted in a similar environment as for P-MLB use-case.

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this magazine.

| Reference | Scheme Name | Input | Machine Learning Tool Used | Output | Result |
|---|---|---|---|---|---|
| [9] | Spatiotemporal Mobility Prediction | Time stamp, Serving Cell, HO occasion | Semi Markov Chain | MPM | More than 80% accurate user future location prediction |
| [11] | Predictive Handover | HO call Data Records BS density, UE velocity | Discrete Time Markov Chain | P-HO | Predictive HO to reduce HO signaling (by 50%-80%) and HO latency (by 49%-60%) |
| [12] | OPERA | Time stamp, Serving Cell, HO occasion, BS density, cell bandwidth | Semi Markov Chain | P-MLB | Proactive load balancing results in 99% satisfied UEs |
| [13] | AURORA | Time stamp, Serving Cell, HO occasion, BS density, BS location | Semi Markov Chain | P-ES | Energy Reduction Gain of about 68% for high traffic demand scenario in UDMN as compared to Always ON approach. |

*Table 1. Summary of MPM and Automation (P-AUTO) use-cases.*

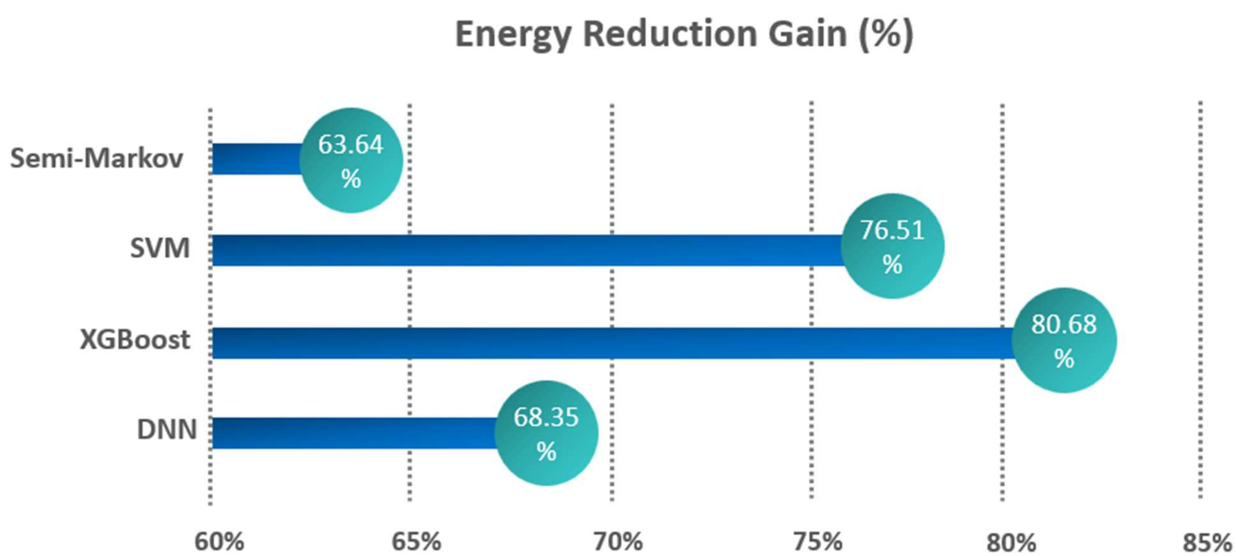


*Fig. 5. Energy reduction gain achieved with various MPM predictors (System model and Simulation scenario is detailed in [13])*

## V. Leveraging A-MMUF Enabling New Use-Cases or Avoiding Conflict Among Multiple Use-Cases

The discussion in the previous section shows the A-MMUF can enable a variety of use-cases that can transform reactive optimization or no automation to proactive optimization or proactive automation of parameters associated with use-case specific KPIs by leveraging the intelligence harness-able from mobility patterns. Other than the P-HO, P-ES and P-MLB use-cases discussed above, other automation use-cases that benefit from MPM to become proactive by leveraging A-MMUF style operation include Proactive-CCO (P-CCO), Proactive-ICIC (P-ICIC). MPM can

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.

transform CCO to P-CCO by enabling proactive optimization of parameters such as Tilt, Transmission-power and CIO, in anticipation of traffic and mobility patterns. Similarly, through A-MMUF adaptation, ICIC can be transformed to P-ICIC where radio resources in future cells of the users are reserved in advance. Moreover, in P-ICIC the scheduling in different sectors of the cells can be done intelligently to minimize cell-edge interference by taking into account user mobility trajectory (MPM) learned e.g., from past HO traces.

Another novel use-case of A-MMUF is its ability to resolve parametric or objective conflicts that till today limit the full automation of cellular network optimization. For example, ES automation function is known to have conflict with CCO use-cases [14] where a cell can be switched off by ES SON causing conflict to objectives of CCO. Similarly, a CCO can have parametric conflict with the MLB where both automation functions may try change the same parameter such as transmission-power or CIO for different objectives. A-MMUF, thanks to its MPM powered ability to predict future cell loads, user location and next cells of users has a potential to enable concurrent operation of multiple automation functions without conflict. The arrival of users in cell switched off by ES automation function can be predicted through MPM to turn it ON preemptively thereby avoiding conflict with CCO.

In light of above discussion, it can be concluded that A-MMUF, in contrast to the state-of-the-art conflict prone reactive automation (SON) can enable multiple automation use-cases to co-exist while ensuring improvement in one KPI does not have negative implication on another associated KPI. While promising in principle, thorough quantitative investigation of the conflict avoidance potential of A-MMUF requires dedicated modeling and analysis for each distinct set of conflicting automation functions and can be subject of future studies.

## CONCLUSION

While the mobility management paradigm has evolved over the past decades from 2G to 5G, we explain why the current mobility management, will not be suitable for the UDMNs. We have highlighted the challenges that stem mainly due to the peculiar characteristics of mmWave, small footprint of small-cells and the intrinsically reactive design and operation of mobility management. To address highlighted challenges, we have a framework for transforming mobility management from a reactive to proactive paradigm by empowering it with machine learning based mobility and traffic prediction models. We propose A-MMUF framework that leverages regular network measurements to predict mobility patterns, next candidate cells for HO, and the time of HO as well as future cell loads. The framework not only solves the identified HO related problem anticipated in UDMNs, but it can also transform mobility from being a challenge to advantage i.e., the knowledge gained from mobility/HO patterns can be leveraged to devise new proactive automation solutions. A-MMUF in turn allows mobility management and conventional automation functions to be proactive rather than reactive meeting the diverse and acute 5G/6G requirements including extremely low latency in URLLC use-case. We also demonstrate the gains of the A-MMUF through three different case studies. The case studies on proactive HO, proactive MLB and proactive ES demonstrates how the presented framework can transform mobility from a bane into a blessing by yielding significant gain in network KPIs. However, observed gains heavily depend on not only the accuracy but also the training time and interference time of underlying mobility prediction models.

## ACKNOWLEDGMENT

This work is supported by the National Science Foundation under Grant Number 1718956 and Qatar National Research Fund (QNRF) under Grant No. NPRP12-S 0311-190302. The statements made herein are solely the responsibility of the authors. For more details about these projects please visit: http://www.ai4networks.com.

## REFERENCES

[1] Ericsson, "Ericsson Mobility Report, 2017-2027," 2021.

[2] C. Pan, R. Liu and G. Yu, "Joint User Association and Resource Allocation for mmWave Communication: A Neural Network Approach," *Journal of Communications and Information Networks,* vol. 6, no. 2, pp. 125-133, 2021.

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.

[3] S. M. A. Zaidi, M. Manalastas, H. Farooq and A. Imran, "Mobility Management in Emerging Ultra-Dense Cellular Networks: A Survey, Outlook, and Future Research Directions," *IEEE Access,* 2020.

[4] 3GPP, "3GPP TR 36.839 - Mobility Enhancements in Heterogeneous Networks," 2015.

[5] S. M. A. Zaidi, M. Manalastas, H. Farooq and A. Imran, "SyntheticNET: A 3GPP Compliant Simulator for AI Enabled 5G and Beyond," *IEEE Access,* vol. 8, pp. 82938-82950, 2020.

[6] 3GPP, "3GPP TS 36.133 - Evolved Universal Terrestrial Radio Access (E-UTRA): Requirements for Support of Radio Resource Management," 2022.

[7] K. Yu, J. Yu and A. Dong, "Cooperative Communication and Mobility for Securing URLLC of Future Wireless Networks," *IEEE Transactions on Vehicular Technology,* 2022.

[8] A. Imran, A. Zoha and A.-D. A. , "Challenges in 5G: How to Empower SON with Big Data for Enabling 5G," *IEEE Network,* vol. 28, no. 6, pp. 27-33, 2014.

[9] H. Farooq and A. Imran, "Spatiotemporal Mobility Prediction in Proactive Self-Organizing Cellular Networks," *IEEE Communications Letters,* vol. 21, no. 2, pp. 370-373, 2017.

[10] H. Gebrie, H. Farooq and A. Imran, "What Machine Learning Predictor Performs Best for Mobility Prediction in Cellular Networks?," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, Shanghai, 2019.

[11] A. M. et al., "Predictive and Core-Network Efficient RRC Signalling for Active State Handover in RANs With Control/Data Separation," *IEEE Transactions on Wireless Communications,* vol. 16, no. 3, pp. 1423-1436, 2017.

[12] H. Farooq, A. Asghar and A. Imran, "Mobility Prediction Based Proactive Dynamic Network Orchestration for Load Balancing With QoS Constraint (OPERA)," *IEEE Transaction on Vehicular Technology,* vol. 69, no. 3, pp. 3370-3383, 2020.

[13] H. Farooq, A. Asghar and A. Imran, "Mobility Prediction Based Autonomous Proactive Energy Saving (AURORA) Framework for Emerging Ultra-Dense Networks," *IEEE Transactions on Green Communications and Networking,* 2018.

[14] H. Y. Lateef, A. Imran, M. A. Imran, L. Giupponi and M. Dohler, "LTE-advanced self-organizing network conflicts and coordination algorithms," *IEEE Wireless Communications,* vol. 22, no. 3, pp. 108-117, 2015.

[15] 3GPP, "3GPP 37.827 - Rel-17 Power Class 2 for EN-DC with LTE band + NR DL with LTE+(TDD) NR UL band," 2022.


Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on May 24,2023 at 23:40:36 UTC from IEEE Xplore. Restrictions apply.