Key Challenges of Mobility Management and Handover Process In 5G HetNets

Sultan Alotaibi
srotaibi@uqu.edu.sa
College of Computer and Information Systems, Umm Al-Qura University, Makkah, 24381 Saudi Arabia

Summary
Wireless access technologies are emerging to enable high data rates for mobile users and novel applications that encompass both human and machine-type interactions. An essential approach to meet the rising demands on network capacity and offer high coverage for wireless users on upcoming fifth generation (5G) networks is heterogeneous networks (HetNets), which are generated by combining the installation of macro cells with a large number of densely distributed small cells Deployment in 5G architecture has several issues because to the rising complexity of network topology in 5G HetNets with many distinct base station types. Aside from the numerous benefits that dense small cell deployment delivers, it also introduces key mobility management issues such as frequent handover (HO), failures, delays and ping-pong HO. This article investigates 5G HetNet mobility management in terms of radio resource control. This article also discusses the key challenges for 5G mobility management.
Keywords:
Mobility management, handover, 5G, HetNets.

1. Introduction

The deployment of a high number of small cells into a HetNet results in a rise in the number of cell edges and an increase in the amount of intercellular contact. This leads to additional handovers HOs and radio link failures [1] [2]. Ping-pong HO (PPHO) occurs when a user connects to an adjacent cell for a brief period before returning to the original source cell. PPHO is undesirable since it wastes energy. Mutual signaling during HO failure (HOF) and PPHO causes additional signal burden, reducing battery life of user devices.

5G mobility management will offer new capabilities that improve the user experience and support future use cases. The core needs of 5G mobility management include uninterruptible HO operation, adjustable HO security, and link monitoring. Moreover, mobility management must design service-specific mobility configurations for multi-slicing networks. Network slicing is a virtual network design that creates numerous virtual networks on top of a physical infrastructure. Apps, services, devices, users, and operators have their own virtual networks. Rather of wasting complicated and costly hardware resources, they directly target devices that need unique functionality, saving time and resources. Because of their asynchronous connection with high-reliability IP telephony and internet access services, internet of things (IoT) and machine-to-machine (M2M) services are allotted distinct network slices. In contrast, the paging process needed to locate the UE, the registration process used to connect the UE to the network and access its services and applications, and mutual signaling during HO processing create network delay and power consumption. Thus, in 5G mobility management, a flexible balance between power saving and delay is required. R radio resource control (RRC) manages communication between UEs and BS. Long-term evaluation (LTE) introduced idle and connected RRC states, whereas 5G brought inactive state lately. There are two RRC states: idle and connected. Inactive is a new state between idle and connected.

Many research work has been done to address management issues in HetNets. The main reason for performing HO is user mobility. In [3], authors developed a minimal variance-neutral approach to estimate user velocity from the number of HOs in HetNets. According to [4] users with frequent HOs are grouped as fast-moving or ping-pong users to minimize total HOs and boost network efficiency in ultra-dense HetNets. This algorithm associates speedy users with macro BSs, whereas ping pong users strive to limit the consequences of unneeded transfers by optimizing their HO settings. Using HetNet as an example, the authors of [5] attempted to optimize cross-tier HO in terms of latency characteristics. Article [6] suggested a network architecture for effortlessly switching between two nearby macro evolved NodeBs (eNB) in the 5G control/user plane split, which was implemented in a prototype. The proposed architecture allows for continuous transmission by combining HO aided micro base station with DC communication. Using the analytical hierarchy processing (AHP) approach, a method is presented in [7] for achieving the weight of the HO metrics and then sorting the cells to pick the optimal HO target using the grey rational analysis (GRA) method. In [8], the authors investigated the implications of channel fading on the management of mobility in HetNet. The findings indicate that raising the sample duration of the HO decision reduces the fading impact while increasing the ping-pong effect. The work in

Manuscript received April 5, 2022
Manuscript revised April 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.4.18

[9] offered a user association rule that was mobility sensitive. The rule attempts to alleviate congestion by sending UEs to tiny BS while simultaneously monitoring dynamic changes in channel conditions induced by user movement in the network topology, which is a common occurrence in networks. It prevents HO and PPHOs between small BS. However, it takes into consideration the unique characteristics of millimeter-wave (mmWave) communication, such as its distinctive directional nature, sensitivity to clogging, and non-line-of-sight propagation propagation effects, and the UE is spread across the network in accordance with these characteristics. Because of their high-frequency nature, the mmWaves are more likely to be obstructed by objects. In order to tackle this challenge, a framework is provided in [10] that anticipates obstacle-caused data rate deterioration before the degradation happens by increasing the status area over time as a result of the succeeding camera pictures. The use of deep reinforcement learning to determine HO timings enabled the researchers to overcome the difficulties associated with dealing with huge dimensional data.

In this paper, key issues, which are significant in handover process, are explained and introduced. In the next section, the important of the power consumption challenge is provided. Third section includes the issues related of using mm-Wave technology. Fourth section sheds the light on the impact of signaling during the process of handover. Also, the security issues are discussed in the fifth section. In the sixth section, the challenge of preserving the energy in the network is discussed. Next, the effect of the load on the process of handover is described. The last section provides a conclusion of this work.

2. Power Consumption

HetNet is comprised of a massive number of BSs that operate on a range of different frequencies. The UE collects measurements from a variety of BSs at both the inter-frequency and intra-frequency levels. If the source cell and the target cell to which HO will occur are both operating at the same carrier frequency, this is referred to as intra-frequency HO. The UE is not required to change the carrier frequency, and measurements are taken in the order in which the cells are received. If the carrier frequencies of the source and target BSs are different, this is referred to as Inter-frequency HO, and the UE is required to change the carrier frequency within the measurement ranges in accordance with the carrier frequencies of surrounding BSs in this case. In this scenario, measurements are carried out in accordance with the quality of the carrier frequencies used. UE target cell selection is accomplished via the use of both intra-frequency and inter-frequency measurements, which are carried out in the order of priority granted by BS. A rise in battery power consumption happens as a result of these observations, and the increase is proportional to the network density.

In addition to causing excessive use of resources and power consumption due to the dense deployment of small cells in HetNet, several issues that reduce QoS, such as interference, frequent and inappropriate HO, HOF, and PPHOs, will also result from the dense deployment of small cells in HetNet. However, HetNet has more frequent high-level operations, the frequency of exchanging mutual signal packets between the source cell, the destination cell, and the user increasing the amount of power used by the network. In mmWave systems, power consumption on mobile devices is a major challenge that need to be properly handled. Because of the nature of mmWave transmissions, it suffers from isotropic path loss in this case. Because of this, mm-Wave systems broadcast with narrow and electrically steerable beams to cope with this problem. To achieve beam steering, a vast number of antennas are utilized. When using analog beamforming, the beams are shaped along a single RF chain, which is shared by all antenna components. This process is carried out in the analog domain, and it is only capable of transmitting and receiving in one way at a time. It also preserves power since it only requires a single analog-to-digital converter, rather than several converters. A single analog-to-digital converter is used in this case, which reduces the amount of power required. However, since it can only broadcast and receive in one way, the system's functionality is limited. In digital beamforming, each antenna element has its own RF chain and data converters, which must be connected in series. It is possible to interpret incoming signals digitally and to guide beams from a receiver/transmitter in a limitless number of directions, allowing for substantially better search times. However, digital systems have a high-power consumption since each antenna requires a separate RF chain and an analog-to-digital converter. Article [11] investigates the power consumption of multi-carrier receivers for systems running at two different frequencies: 28 GHz and 140 GHz. In the case of the RF front-end components, it has been discovered that the mixer and phase shifter absorb the majority of the available power. In contrast, the hybrid analog-digital design used in the transceiver is regarded to be a more cost-effective approach. On the other hand, the hybrid analog-digital design used in the receiver/transmitter is supposed to be more cost-effective alternative. In comparison to digital beamforming, hybrid beamforming makes use of fewer RF chains, allowing for the utilization of more antenna array components while lowering energy consumption and system design complexity.

3. Using mm-WAVE

Due to the dramatic increase in mobile traffic, mmWave offers a critical opportunity to resolve the problem between capacity demands and the scarcity of available spectrum. To boost communication capacity, the mmWave utilizes the huge spectrum of the mmWave. Due to low order modulation, mmWave bands range between 30 to 300 GHz, unlike the congested sub-6GHz frequencies. mmWave has a lot of advantages, but it also has a lot of drawbacks.
Increased transmission losses and lower signal levels are caused by precipitation's effects on radio waves absorption, scattering, and diffraction. If this happens, the propagation of mmWave signals would suffer greatly and the signal attenuation will be quite severe. The attenuation induced by rain rises dramatically as the frequency, intensity, and effective duration of the rain increase. As a result, the communication link's reliability may be degraded, and any existing connections may become unreachable. The design of 5G system channels requires the use of actual measurement data in order to provide more accurate estimations and improved performance. According to [45] r ain attenuation of 38 GHz was shown to be crucial, with an 18.4 dB/km loss potential.

Fast channel fluctuations and significant free space loss and atmospheric absorption are major issues for mmWave connectivity. Both the UE and the BS employ progressive array beamforming and massive MIMO approaches to address this issue and boost system capacity. Beamforming methods may be used to make up for the propagation loss in directive communication while also increasing the system capacity. The interband interference is considerably reduced by using beamforming methods, which reduce the performance requirements for each antenna and RF circuit. MIMO architecture and beamforming techniques in both the UE and the BS offer significant energy efficiency as well as increased communication reliability.

mmWave signals encounter issues including significant path loss, severe channel interruption and blockage caused by construction materials like bricks and mortar and even the human body [46]. Consequently, the quality of UE-serving cell connectivity is very changeable owing to variables like as the movement of barriers or the shift of the body's location relative to mobile terminal, which may cause fast reductions in signal strength. A UE may be linked to many cells at the same time and have various signal channels thanks to multiple connectivity. A data path may be changed if connection quality drops, thereby ensuring the continuation of communication by preventing the quality of the connection from deteriorating. Multiple interconnectivity between 5G mmWave cells or 5G mmWave cells and 4G cells is possible in cellular mmWave networks. Multiple 5G mmWave cell connectivity offers increased bandwidth, while 5G mmWave cell connectivity with 4G cells provides more robust communication. [47] proposes a multicellular measurement reporting system. To examine the angular space, each UE sends SRS in various directions over time. Each prospective service cell in the system searches in all directions to better understand the channel dynamics and the SRS received from UEs. A central controller then gathers complete directional information from candidate cells, selecting and timing serving cells. On the basis of simulation findings, the suggested system can create digital beams in BS much faster and can observe many mmWave cells at a tolerable cell density. [48] provides a DC protocol that enables UEs to connect to both 4G and 5G cells. An uplink control signaling system with a local coordinator allows for quick patch switching in case of failure. In the PDCP layer, the local coordinator handles inter-cellular traffic and performs control plane functions like path switching. The DC framework provides smoother mobility management than the hard HO diagram. In uncertain conditions, dynamic TTT adaptation may help enhance critical decision scheduling. The simulation results showed that the suggested framework improved efficiency stability and improved metrics such as latency, packet loss, and control signal load. However, with mmWave, multiple interconnectivity is more challenging due to omnidirectional broadcasts. In order to offer a good service, every possible link between the network and the UE must be continually evaluated. A change in course is also problematic since the UE and BS can only listen to one way at a time. Instead of continually monitoring each link, other approaches should be implemented. During cell finding procedure, the beam scans the angular coverage zone for synchronization signals. At this point, the best performing beam pair between UE and gNB is determined. Sequential search generates high access latency and poor starting efficiency. So we provide a beam sweeping model based on the dynamic distribution of user traffic using repeated neural networks (RNN). Data from cellular network call detail records (CDR) offer spatial distributions of users. mmWave employs RNN to accurately forecast CDRs to efficiently detect scan direction in the cellular system.
5G mmWave systems should have small RF receiver and transmitter to save energy. As previously stated, hybrid analog/digital designs are employed to minimize RF chains. These designs make scalable array systems simpler. The digital pre-encoder processes data flows before converting them to RF frequency and mapping them to all antennas for transmission across a phase shifter network. The construction of analog and digital beamforming matrices [50] is the main problem in hybrid analog/digital systems. When a UE initially approaches the BS's network coverage, the beams between the UE and BS must be aligned. This process places a heavy demand on mmWave systems utilizing hybrid beamforming. Reducing new user discovery time is critical. Also, the user and BS locations

are unknown, therefore the search operation introduces overheads. The literature suggests many methods for locating users. Arrival data are dependent on user entrance points, routes, and buildings. [51] proposes an online technique to acquire these information b y scanning more often on the locations where users spend the most time. These data may also be used to improve a previously failing feature. So performance is evaluated using actual data. The offered approach is claimed to greatly decrease discovery time compared to non-statistical methods. Article [52] proposes a gradient descent approach for 5G mmWave beam alignment, allowing users to characterize pure Nash equilibrium and find ideal beam widths. The game is defined by parameters like beam width, distance between transceivers, alignment ability, and transmit power. Also, . also, [53] proposes a coordinated machine learning and beamforming solution for mobile mmWave applications. The coordinated beamforming method effectively supports several BS users. The deep learning model predicts omni or semi-omni beam models from signals received from scattered BSs. Signals from distributed BS reveal information needed to adjust to the user's setting. The user using a deep learning model and a robust beamforming system was effective in responding to changing conditions.

4. Signaling

HetNet 5G's massive deployment of small cells has increased the frequency of inter-cell HO. Inter-cell communication is required for users to be registered with target cells in HO mode by exchanging signal packets between source cells, target cells, and the user's device. Increasing the frequency of HO will increase the signal strain on the network, resulting in more data transmission disruptions. This results in a trade-off between the added signal burden from the frequent HO and the network's overall coverage. Because of this, effective mobility management solutions are vital in achieving this balance. Lower the number of HOs in order to reduce the signal burden. Aricle [64] proposes a method for adaptively adjusting the HO margin and TTT based on the user's speed and the received signal's intensity as a reference. Consequently, the goal is to reduce the HO number and the HOF rate. The suggested technique greatly reduces the average PPHO and HOF parameters compared to research in the literature. In addition, it lowers communication delay and interruption. In [65], a wireless test-bed intended to assess the performance of ultra-dense networks and the mobility methods associated with these networks is suggested. The UE's position may be tracked in advance using a mobility management system, ensuring that users are provided uninterrupted HO service. To determine exactly where it is inside a network, the UE transmits short-range signals (SRSs). Thus, it reduces the need for HO, but it also adds complexity to the system. Signal burden is evaluated in terms of network density, user mobility, and session attributes by means of the suggested signaling strategy in [66]. The HO signal latency has also been reduced to the greatest extent possible using a clever HO strategy. Mobility signals are reduced when this approach is used in networks that use a control/data separation architecture (CDA). HO delay and signal load are also reduced in simulation findings.

5. Security

Mobility security in 5G HetNet, which comprises of the deployment of a large amount of small cells from various technologies, is a significant issue due to the frequent HO. Malicious users may take precautions against network effects such as Man-in-the-Middle attacks, Denial of Service attacks, impersonation attacks, and repetition attacks by using mutual authentication between UEs and BS. A secure transport authentication system is essential in order to take safeguards against these threats and to ensure reliable communication while moving data across various networks. Due to the messaging involved in the authentication method, as well as the interfacing of the BSs, it is possible that HO delays may arise. Therefore, launching an authentication process for each HO is not a cost-effective solution. In [73], authors propose a more secure vertical HO authentication system that is symmetric with the key distribution extensible authentication protocol-transport layer security (EAP-TLS) in order to guarantee that the user equipment obtains certificates from a foreign network. EAP is an authentication mechanism that is often used in network and internet connections, and EAP-TLS is a standard that is maintained amongst wireless providers that uses the TLS protocol to authenticate users. In [74] t he anonymous mutual authentication by key agreement was presented and it took use of the trapdoor collision feature of the chameleon hash functions as well as the tamper resistance of blockchains. For the mobile relay node, two fixed trajectory groups for the HO node authentication mechanism are provided in [75]. This ensures that MRNs in a train arrive at the next BS before authentication can be done and that communication continues uninterrupted.There is also an increase in the number of paging processes that must be performed in order to locate a user in the network, and this results in an overall signaling load increase. Paper [67] proposes a novel gNB-based UE mobility tracking system that does away with TAL and the accompanying IoT/UE monitoring TAU/RAN-based notification region. Power savings will be realized since IoT/UEs do not communicate with the TAU. The gNBs take over the monitoring and placement of the IoT/UEs by providing them with their permanent placements. Because the paging technique has been abolished, these devices may be accessed more quickly and easily. It is also worth noting that this technique adds some signalling burden to the network with a very short paging latency. Compared to standard TAU and paging designs, simulation results indicated a 92% decrease in signaling burden.

6. Efficiency of the Energy

High-capacity HetNets are a viable approach for meeting the large mobile traffic demand and increasing capacity requirements in 5G networks. Small base stations, which are deployed in large numbers and incorporated into HetNets, may handle these demands while increasing energy usage. The reason for this is that small cells have energy expenditures even in cases when there are no users linked with them. This situation necessitates either moving the load from cells with high load to these idle cells or shutting down the cells if the system capacity will not be adversely affected in any significant way. The overall amount of energy used by BSs throughout the whole network is significant. Network operators are thus attempting to develop effective power management techniques in order to cut operating expenses (OPEX). Cost models that combine technical and commercial operations, as well as administrative expenditures are known as operating expenses (OPEX) [69]. Additionally, a large amount of energy is used during the movement of users to and from other cell sites. It is critical to perform accurate energy efficiency estimates during the decision phase of the HO process in order to prevent excessive energy consumption due to poor connection conditions of the target cell. As a result of the energy efficiency transfer techniques that will be developed, the battery life of users may be significantly increased. However, attempts to minimize energy usage are becoming more crucial because to the rise in the carbon footprint caused by the mobile communication sector and expanding at an exponential rate over time. As described in [70], the authors provide a fuzzy logic-based game theoretical framework for optimizing the optimal transmission BS power levels to serve UEs in order to maintain energy efficiency in HetNets. The fuzzy HO system that has been suggested is composed of two modules: the HO decision and the target BS selection modules. In the HO decision, the network operator may flexibly modify the intended energy efficiency, PP rate, and efficiency by considering a variety of characteristics such as the speed of the network, the SINR, the efficiency, and the BS load. According to the order of preference generated by taking into account all of the characteristics, the target BS selection module determines which BS is most appropriate to use. Ping-pong transfers for high-speed users may be managed more efficiently, according to the findings of simulations, which have shown that energy usage can be reduced. In [71], a HO algorithm is developed that allows the system to self-optimize in order to enhance energy efficiency and the PPHO ratio of the system. System energy reduction gain (ERG) and PPHO rate threshold parameters are selected in each time period using the sample of power consumption and SINR performance, and the network compares the ERG and PPHO rate conditions with two threshold values. A metric known as ERG is a percentage-based measure of the difference in energy consumption between two systems. System performance is improved as a consequence of the comparison, and the following time period's TTT and HOM values are optimized as well. Because of their high mobility, unmanned aerial vehicles (UAVs) allow for more flexible and faster distribution of communication systems, as well as higher connection capacity because they have longer line of sight channels than ground-to-ground connections, compared to traditional ground-to-ground connections. UAVs are expected to play a significant part in future wireless networks as a result of these characteristics. For a UAV-enabled mobility transition system, the study in [83] attempts to investigate spectrum efficiency maximization and energy efficiency maximization designs and expose essential trade-offs by optimizing communication time allocation and UAV trajectory. The results of simulations have shown that there is a new fundamental trade-off between spectrum efficiency maximization and energy efficiency maximization in 5G wireless communication systems. Ultra-reliable low latency communications (URLLC) is a service designed to be employed in essential applications demanding low latency in 5G wireless communications systems. The work in [72] offers a HO approach for 5G URLLC that bypasses the function of the source gNB, with the goal of sending HO requests directly between the UE and the target gNB based on the measurements taken by the UE. As a result, by lowering the user plane delay, it is possible to enhance energy efficiency while also achieving HO more quickly.

7. Balancing The Load

HetNet architectures, in which small base stations collaborate with large base stations (BSs), enable network coverage to be expanded at a reduced cost while simultaneously enhancing network performance characteristics such as spectrum efficiency and energy efficiency. Because of the random arrangement of the cells and the mobility of the UEs in very dense HetNets, there is a load imbalance between the cells in these networks. The rate of HOF grows when there is a load imbalance within the network, and the efficiency of network performance decreases. When using traditional user association architectures that determine the target cell for which HO will be recognized using the maximum signal-to-noise ratio (max-SINR) criterion, most users will associate with macro BSs because the power supplied by the macro BSs is greater than the power supplied by the small BSs. As a result, macro BS resources become overburdened while small base stations resources are underutilized. Consequently, there is a load imbalance between the BSs in the network, and the

QoS given to users is lowered. It is theoretically possible that handing over a UE to an already overloaded cell may result in a scarcity of resources in the loaded cell as well as a reduction in the QoS offered to both current UEs in the cell and the new UE, with the consequence that the resultant HO will be regarded as failed. Therefore, current research has given significant attention to load balancing, which will be accomplished by moving the load from overcrowded cells to idle cells. According to the load circumstances of small cells in the environment, the parameters to be employed in the HO of UEs are altered in these experiments, which results in improved performance. A thing to emphasize is that the parameters must be correctly adjusted; otherwise, inefficient use of network resources and insufficient QoS will be delivered to UEs would arise as a consequence of this.

In [54], a cluster-based load balancing algorithm is presented in lieu of a mobility load balancing method throughout the whole network to reduce load imbalance and improve network performance by eliminating load unbalance. To generate clusters from overloaded cells and their n-layered neighbors, the proposed approach models the network in order to find the shortest path between them. Because load balancing is done locally by dynamically altering cell individual off-set (CIO) settings for each cluster, load balancing is avoided when unneeded load balancing operations are taken. Compared to a network that does not use load-balancing techniques, simulation findings reveal that the suggested method results in a 6.42 percent improvement in network performance throughput in the case of low-speed user equipment (UE). In general, current load balancing systems describe this issue in terms of the complete CSI between the user and each BS, and this formulation at the global level complicates structures while also creating extra signal load. A low complexity sequential offloading technique is developed in [55] to tackle the load balancing issue. Instead of a global CSI feedback structure, each user detected and reported sync signal levels of neighboring BSs. As a consequence, it has been proved via simulation results that it is possible to get outcomes that are near to ideal load balancing performance while maintaining a reduced level of complexityIn [56], a utility-based mobility load balancing (UMLB) method is described that detects the edge UEs of an overloaded cell that are necessary for load transfer to occur from an overloaded cell to lightly loaded surrounding cells and offers the optimal nearby cell HO by calculating the total utility of these candidate users with each adjacent cell. Additionally, it introduces a load balancing efficiency factor (LBEF), which is a newly added term that will indicate the order of overloaded cells. According to simulation findings, UMLB reduces standard deviation while maintaining a higher average UE data rate than existing load balancing techniques.

It is anticipated that the adoption of machine learning algorithms in load balancing management would improve the service provided to users and bring about significant benefits. in [57], authors present a deep learning-based mobility strength optimization approach that learns the right values of the parameters necessary for each cell's mobility model while it is being deployed in a network. The optimal mobility setting for transmission parameters has been determined based on the distribution of users and the pace at which they move across the network. A mobility-sensitive load balancing strategy has been implemented in order to offer each user with about the same Quality of Service (QoS). HO performance optimization and learning parameter selections are shown by simulation results, which demonstrate that the proposed system outperforms the benchmarked reference system. It is possible to integrate real-world data urban incident detection with proactive load balancing using the framework provided by [58]. First, it was suggested that urban event detection be used to forecast changes in cellular hot spots based on data from Twitter. Then, using a proactive load balancing method, simulation is carried out while taking into account the distorted hot spots in the metropolitan area. The last step involves optimizing the proactive load balancing method in order to determine the optimal activation time.

BSs assisted by unmanned aerial vehicles (UAVs) are a promising solution to the issue of balanced network load since they have qualities such as flexible distribution, wireless coverage everywhere, and benefits in terms of data rate that are difficult to match. However, determining how unmanned aerial vehicles (UAVs) can be deployed independently and dynamically inside a network is a significant difficulty. UAV BS intelligent distribution strategy based on machine learning is suggested in [59], and performance evaluations are carried out on an actual data set to assess its effectiveness. Data preprocessing is carried out in order to process data received over the network, clear it, and convert it to another format. Hybrid predications, which contain the Autoregressive integrated Moving Average (ARIMA) model, as well as an extreme gradient-boost model, are then created using the hybrid technique. The ARIMA prediction model is a linear prediction model, while the XGBoost prediction model is a nonlinear prediction model. It is used to estimate the number of future users using XGBoost nonlinear prediction based on ARIMA's linear prediction, which is a part of the suggested hybrid model. According to the predictions made in the previous step, unmanned aerial vehicles (UAVs) are deployed in hot spot locations to dynamically satisfy the needs. The results of the simulation have shown that this strategy is an effective scheme for load balancing.

8. Conclusion

In this article, a detailed discussion of the difficulties faced in 5G HetNets mobility management is provided, as well as a description of the solutions proposed in the literature in order to cope with these difficulties. Six key challenges for mobility management in 5G HetNets are described in this work. First, the important of the power consumption challenge is provided. Then, the issues related of using mm-Wave technology and its impact on the process of handover is discussed. Also, this work sheds the light on the impact of signaling during the process of handover. Additionally, the security issues are discussed with its significant consideration. Moreover, the challenge of preserving the energy in the network is discussed and explained. In addition, the effect of the load on the process of handover is described.

References

[1] Zhang, H., Meng, N., Liu, Y., & Zhang, X. (2016). Performance evaluation for local anchor-based dual connectivity in 5G user-centric network. *IEEE Access*, *4*, 5721-5729.

[2] Tufail, A., Namoun, A., Alrehaili, A., & Ali, A. (2021). A Survey on 5G Enabled Multi-Access Edge Computing for Smart Cities: Issues and Future Prospects. *International Journal of Computer Science & Network Security*, *21*(6), 107-118.

[3] Tiwari, R., & Deshmukh, S. (2019). MVU estimate of user velocity via gamma distributed handover count in HetNets. *IEEE Communications Letters*, *23*(3), 482-485.

[4] Hasan, M. M., Kwon, S., & Oh, S. (2018). Frequent-handover mitigation in ultra-dense heterogeneous networks. *IEEE Transactions on Vehicular Technology*, *68*(1), 1035-1040.

[5] Xu, X., Tang, X., Sun, Z., Tao, X., & Zhang, P. (2019). Delay-oriented cross-tier handover optimization in ultra-dense heterogeneous networks. *IEEE Access*, *7*, 21769-21776.

[6] Zhang, Z., Junhui, Z., Ni, S., & Gong, Y. (2019). A seamless handover scheme with assisted eNB for 5G C/U plane split heterogeneous network. *IEEE Access*, *7*, 164256-164264.

[7] Alhabo, M., Zhang, L., & Nawaz, N. (2019). GRA-based handover for dense small cells heterogeneous networks. *IET Communications*, *13*(13), 1928-1935.

[8] Vasudeva, K., Şimsek, M., López-Pérez, D., & Güvenç, I. (2015, June). Impact of channel fading on mobility management in heterogeneous networks. In *2015 IEEE International Conference on Communication Workshop (ICCW)* (pp. 2206-2211). IEEE.

[9] Cacciapuoti, A. S. (2017). Mobility-aware user association for 5G mmWave networks. *IEEE Access*, *5*, 21497-21507.

[10] Koda, Y., Nakashima, K., Yamamoto, K., Nishio, T., & Morikura, M. (2019). Handover management for mmwave networks with proactive performance prediction using camera images and deep reinforcement learning. *IEEE Transactions on Cognitive Communications and Networking*, *6*(2), 802-816.

[11] Skrimponis, P., Dutta, S., Mezzavilla, M., Rangan, S., Mirfarshbafan, S. H., Studer, C., ... & Rodwell, M. (2020, March). Power consumption analysis for mobile mmWave and sub-THz receivers. In *2020 2nd 6G Wireless Summit (6G SUMMIT)* (pp. 1-5). IEEE.

[12] Shayea, I., Abd. Rahman, T., Hadri Azmi, M., & Arsad, A. (2018). Rain attenuation of millimetre wave above 10 GHz for terrestrial links in tropical regions. *Transactions on Emerging Telecommunications Technologies*, *29*(8), e3450.

[13] Lu, J. S., Steinbach, D., Cabrol, P., & Pietraski, P. (2012). Modeling human blockers in millimeter wave radio links. *ZTE communications*, *10*(4), 23-28.

[14] Giordani, M., Mezzavilla, M., Rangan, S., & Zorzi, M. (2016, June). Multi-connectivity in 5G mmWave cellular networks. In *2016 Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)* (pp. 1-7). IEEE.

[15] Polese, M., Giordani, M., Mezzavilla, M., Rangan, S., & Zorzi, M. (2017). Improved handover through dual connectivity in 5G mmWave mobile networks. *IEEE Journal on Selected Areas in Communications*, *35*(9), 2069-2084.

[16] Li, L., Wang, D., Niu, X., Chai, Y., Chen, L., He, L., ... & You, X. (2018). mmWave communications for 5G: implementation challenges and advances. *Science China Information Sciences*, *61*(2), 1-19.

[17] Soleimani, H., Parada, R., Tomasin, S., & Zorzi, M. (2019). Fast initial access for mmWave 5G systems with hybrid beamforming using online statistics learning. *IEEE Communications Magazine*, *57*(9), 132-137.

[18] Attaoui, W., Bouraqia, K., Sabir, E., Benjillali, M., & Elazouzi, R. (2019, June). Beam alignment game for self-organized mmWave-empowered 5G initial access. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)* (pp. 2050-2057). IEEE.

[19] Alkhateeb, A., Alex, S., Varkey, P., Li, Y., Qu, Q., & Tujkovic, D. (2018). Deep learning coordinated beamforming for highly-mobile millimeter wave systems. *IEEE Access*, *6*, 37328-37348.

[20] Gures, E., Shayea, I., Alhammadi, A., Ergen, M., & Mohamad, H. (2020). A comprehensive survey on mobility management in 5G heterogeneous networks: Architectures, challenges and solutions. *IEEE Access*, *8*, 195883-195913.

[21] Malm, N., Zhou, L., Menta, E., Ruttik, K., Jäntti, R., Tirkkonen, O., ... & Leppänen, K. (2018, July). User localization enabled ultra-dense network testbed. In *2018 IEEE 5G World Forum (5GWF)* (pp. 405-409). IEEE.

[22] Mohamed, A., Onireti, O., Imran, M. A., Imran, A., & Tafazolli, R. (2016). Predictive and core-network efficient RRC signalling for active state handover in RANs with

control/data separation. *IEEE Transactions on Wireless Communications*, *16*(3), 1423-1436.
[23] Huang, J., & Qian, Y. (2020). A secure and efficient handover authentication and key management protocol for 5G networks. *Journal of Communications and Information Networks*, *5*(1), 40-49.
[24] Zhang, Y., Deng, R. H., Bertino, E., & Zheng, D. (2019). Robust and universal seamless handover authentication in 5G HetNets. *IEEE Transactions on Dependable and Secure Computing*, *18*(2), 858-874.
[25] Ma, R., Cao, J., Feng, D., Li, H., & He, S. (2019). FTGPHA: Fixed-trajectory group pre-handover authentication mechanism for mobile relays in 5G high-speed rail networks. *IEEE transactions on vehicular technology*, *69*(2), 2126-2140.
[26] Alsaeedy, A. A., & Chong, E. K. (2019). Mobility management for 5G IoT devices: Improving power consumption with lightweight signaling overhead. *IEEE Internet of Things Journal*, *6*(5), 8237-8247.
[27] Verbrugge, S., Pasqualini, S., Westphal, F. J., Jäger, M., Iselt, A., Kirstädter, A., ... & Demeester, P. (2005, February). Modeling operational expenditures for telecom operators. In *Proceedings of Conference on Optical Network Design and Modeling* (pp. 455-466).
[28] Vasudeva, K., Dikmese, S., Güven, İ., Mehbodniya, A., Saad, W., & Adachi, F. (2017). Fuzzy-based game theoretic mobility management for energy efficient operation in HetNets. *IEEE Access*, *5*, 7542-7552.
[29] Zhang, B., Qi, W., & Zhang, J. (2018, January). An energy efficiency and ping-pong handover ratio optimization in two-tier heterogeneous networks. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 532-536). IEEE.
[30] Zhang, J., Zeng, Y., & Zhang, R. (2017, May). Spectrum and energy efficiency maximization in UAV-enabled mobile relaying. In *2017 IEEE International Conference on Communications (ICC)* (pp. 1-6). IEEE.
[31] Mukherjee, A. (2018). Energy efficiency and delay in 5G ultra-reliable low-latency communications system architectures. *IEEE network*, *32*(2), 55-61.
[32] Hasan, M. M., & Kwon, S. (2019). Cluster-based load balancing algorithm for ultra-dense heterogeneous networks. *IEEE Access*, *8*, 2153-2162.
[33] Han, P., Zhou, Z., & Wang, Z. (2020). User association for load balance in heterogeneous networks with limited CSI feedback. *IEEE Communications Letters*, *24*(5), 1095-1099.
[34] Addali, K. M., Melhem, S. Y. B., Khamayseh, Y., Zhang, Z., & Kadoch, M. (2019). Dynamic mobility load balancing for 5G small-cell networks based on utility functions. *IEEE Access*, *7*, 126998-127011.
[35] Mohajer, A., Bavaghar, M., & Farrokhi, H. (2020). Mobility-aware load balancing for reliable self-organization networks: Multi-agent deep reinforcement learning. *Reliability Engineering & System Safety*, *202*, 107056.
[36] Ma, B., Yang, B., Zhu, Y., & Zhang, J. (2020). Context-aware proactive 5G load balancing and optimization for urban areas. *IEEE Access*, *8*, 8405-8417.
[37] Hu, J., Zhang, H., Liu, Y., Li, X., & Ji, H. (2019, April). An intelligent uav deployment scheme for load balance in small cell networks using machine learning. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1-6). IEEE.

Sultan Alotaibi received the B.Sc. degree in Computer Science from Umm Al-Qura University, the M.Sc. and Ph.D. degrees from University of North Texas, USA. He is currently an Assistant Professor with the Department of Information Systems, College of Computers and Information Systems, Umm Al-Qura University, Saudi Arabia. His research interest includes wireless communication networks, LTE, LTE-A, radio resource scheduling, 5G and data quality in wireless networks.