[1]CH. Srilakshmi Prasanna

[2] S. Zahoor-ul-Huq

[3] P. Chenna Reddy

Utilization Based Load Balancing Algorithm For Handover Reduction In 5G Networks

Journal of Electrical Systems

Abstract: - The implementation of a load balancing algorithm involves distributing the load from congested small cells to less burdened ones within the network, taking into account mobility parameters. The traditional load balancing algorithms only considers the edges of the devices (UEs) when it comes to the offloading process. The proposed algorithm considers both the non-edge and edge-UEs for minimizing the load on the small cells. Subsequently, it relocates high-speed UEs to the underutilized Macrocell and slower-moving UEs to nearby small cells. The algorithm relies on categorizing UEs into fast, slow, and very slow groups to make optimal handover decisions. Simulations have demonstrated that this approach results in the fewest handovers and the highest throughput.

Keywords: *Load Balancing, Handover, Mobility, User Environment, Small Cell, Macro Cell*.

I. INTRODUCTION

The lack of resources in small cell networks can cause load-imbalance issues. For instance, if a UE wants to enter a cell that has an oversaturated service area, it will probably cause poor Quality of Service (QoS) because the resources of the other nearby cells will not meet the requirements. Through the adjustment of the load statuses of small cells, MLB can distribute the load among them [1]. Subsequently, the individual offsets for neighboring and serving cells are fine-tuned based on reported measurements, potentially influencing the network's performance. To mitigate these effects and ensure network stability, it is essential to implement appropriate load balancing mechanisms.

A small cell plays a crucial role within the 5G network by enabling it to meet the growing data demand. It is a low-cost and efficient radio-access point [7-8]. Although small cells were originally designed to support the increasing number of macro-cells, Moreover, they can be deployed in close proximity to each other to enhance the wireless network's capacity[9]. This makes them an ideal technology for future networks. The rapid emergence and growth of small cells has led to their availability in both residential and non-residential areas [10]. They can be planned or unplanned and can be installed by the service provider without any configuration or planning requirements. Unlike a traditional macro network, small cells are very low-cost and can be randomly distributed [11]. Unlike a traditional macro network, small cells are very low-cost and can be randomly distributed. This eliminates the need for the service provider to set up a specific site and configuration for each small cell.

The growing population of mobile users (UEs) within a small cell network can result in network load imbalance [12-14]. This concern can have an impact on the network's performance, affecting aspects such as success rate, capacity, handover, and the overall quality of service (QoS). The lack of resources in the network can also lead to poor QoS and increase the likelihood of failure when the incoming users try to enter the network's other cells. To improve the network's performance, the network needs to adopt proper management and configuration.

In order to achieve high performance, system parameters are manually adjusted in existing networks. However, with the rapid evolution of networks, it is becoming harder to perform this type of manual tuning. To simplify the operation, an LTE system incorporated a self-organized network. This method allows the network to automatically configure and optimize itself [15]. The three different types of SON algorithms are distributed, hybrid, and centralized. They are used to improve the performance of small cells. Some of the components of these are mobility load balancing, FHM, and MRO. Through MLB, small cells can receive the load they need to improve their Quality of Service (QoS) and increase system capacity. It uses information collected from the cells to set up optimal boundaries and offload the unloaded units (UEs) [16-17].

[1] *Corresponding author: Research Scholar, CSE Department, JNT University Anantapur (JNTUA), Ananthapuramu, A.P, 515002, India, email: sirikesava999@gmail.com

[2] G. Pulla Reddy Engineering College, Kurnool, Affiliated to Jawaharlal Nehru Technological University Anantapur, Ananthapuramu

[3] JNTUA College of Engineering, Ananthapuramu, Constituent college of Jawaharlal Nehru Technological University Anantapur, Ananthapuramu

One of the most expensive components of the MLB operation is the implementation of load balancing mechanisms. This procedure guarantees that the network satisfies the minimal handover request count [18]. In [19], the authors proposed the utility-based approach which is used to manage the load balancing process on mobile devices. It takes into account the various factors that affect the performance of small cells, such as the speed category of the Edge UEs and the non-edge ones. The main objective of this method is to identify the fastest and slowest Edge UEs to transfer them to either the Macrocell or the neighboring under-loaded ones.

The state of the UE is determined by the terms "fast," "slow," and "very slow." The authors use these to determine the optimal approach for each devices. The estimated time that a UE will stay in a serving cell is also taken into account [20]. The remaining service time RT is defined as the estimated amount of time that a user will spend staying at the neighbour cell. The session time denotes the duration required by the device to utilize a specific application. Conversely, the remaining session time, denoted as RQ, represents the time required for the device to fulfill the remaining workload. If the RQ of the UE is greater than its remaining service time S, then it will be transferred to the Macrocell instead of the small cells. This eliminates the need for them to handover frequently.

The UE is categorized as fast-moving at its edge if its session duration surpasses both the RT and the STN. This implies it will be moved to the macro cell to minimize frequent handovers. Additionally, if the UE's RQ exceeds both the RT and the STN, it is categorized as a slow UE. On the other hand, if the UE's RQ is less than the RT, it is categorized as a very slow UE. This type of UEs are not handed over because of the session of UE is ended before leaves the cell. For instance, in Figure 1, Cell A is loaded with 80%, cell B and Cell C is loaded with 55% and 65%. The Adaptive threshold overload value is taken as 65%. Let us consider that, in Cell A, UE1 and UE4 are fast, UE3 is slow, UE2 is static, UE5 are very slow. When the load balancing algorithm is implemented for Cell A, the candidate UEs are identified as {UE5, UE4, and UE3}, selected based on their RSRP (Reference Signal Received Power) values. For example, if each UE's average load consumption is 5%, the load balancing process entails transferring UE4 and UE5 to Cell B. However, UE3 remains in Cell A due to the overload. After applying the proposed handover algorithm to Cell A, the candidate UEs are identified as {UE1, UE4, UE3}, with UE5 excluded from handover due to its slow movement, avoiding unnecessary handover operations. UE4, being fast-moving, is transferred to the macro cell to minimize frequent handovers between cells. UE1 is also moved to the Macrocell because of its fast movement. Consequently, UE3 is handed over to Cell B due to its slow movement. This results in Cell A and Cell B having loads of 65%, with only three handovers, which is a minimal number compared to the load balancing algorithm's operation.

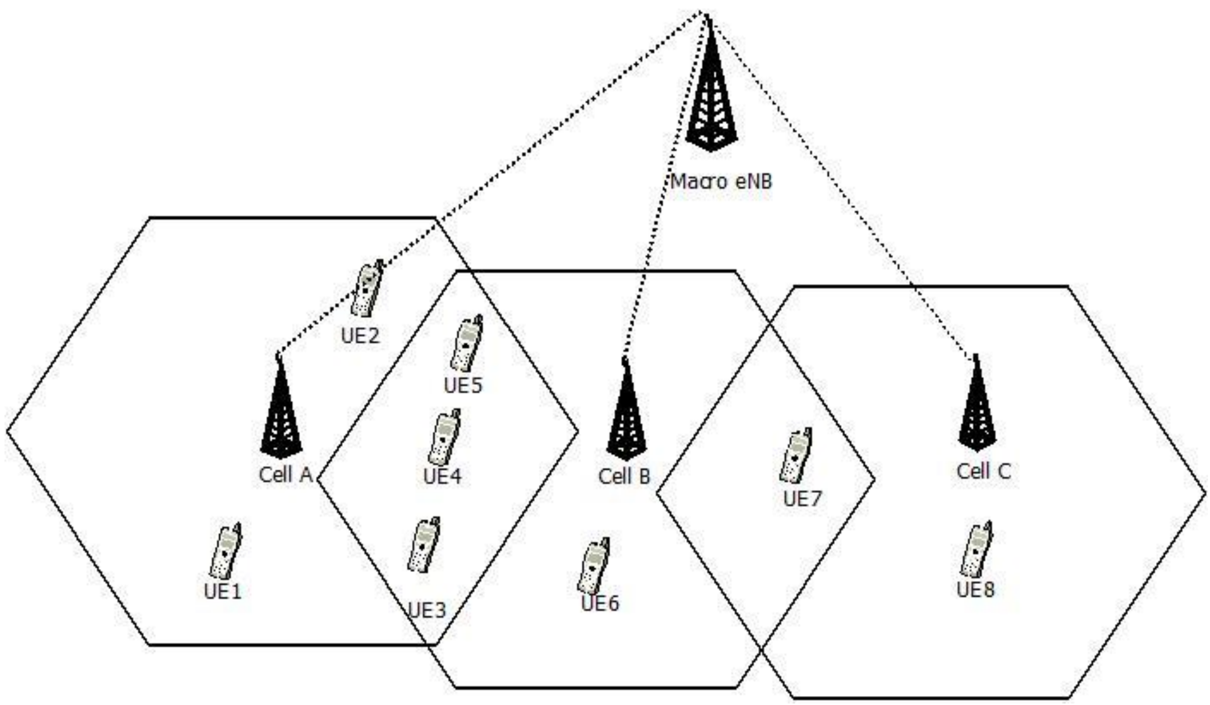

Figure 1: Architecture of Heterogeneous Radio Access Network.

II. SYSTEM MODEL

Let us consider the heterogeneous network which is having the Macrocell M and the small cells w={1, 2, ….W} which is shown in Figure 1. It is having the set of users x={1,2…X}. The core network functionalities, including the Serving Gateway (S-GW), Mobility Management Entity (MME) within the Evolutionary Packet Core (EPC) extended to the small cells via the S1 interface. Each cell in the network is interconnected with the X2 interface, where it provides the handover process directly in the communication [3]. Due to this, the users in the cells can switchover to any cell. The centralized Self organized network (C-SON) is applied to the network where it has the functionality of Operation and Management System (OAM) and evolved Node B (eNBs).

The centralized SONs gathers the information from different small cell C-SONs and updates the information periodically to perform the load balancing operation with less number of handovers.

A. Problem Formulation for Load balancing Mechanism

The primary aim of this research is to enhance QoS and achieve load balancing across the cells. To address this objective, the problem is framed with the aim of reducing the load on small cells by minimizing the frequency of handovers. Therefore, the problem is formulated as:

$$Max\ (QoS\ (v)) = Min\ (\sigma(v), H(v)) \quad (1)$$

$$subject\ \ to\ R1: \sum_{j\in J} B_{w,j}(v) N_{w,j}(v) \leq Q,\ \forall w \quad (2)$$

$$R2: \sum_{j\in J} B_{w,j}(v) = 1\ \forall x \quad (3)$$

Where σ(v) is the standard deviation of the load on the cell, H(V) represents the count of handovers necessary for achieving load balancing within the network. $B_{w,j}(v)$ represents the binary indicator, if $B_{w,j}(v)=1$ then the user j is served by the small cell w. $N_{w,j}(v)$ represents the physical resource blocks which are allocated to the user j by small cell w at time period v. Q represents the overall count of physical resource blocks accessible within the small cell. w. Hence, the Rule R1 denotes that the total number of physical resource blocks of small cell w at time v cannot exceed the cell capacity and Rule R2 denotes that the user environment UE cannot be associated with multiple cells at a time.

III. REMAINING SERVICE TIME AND SESSION TIME ESTIMATION

A.Remaining Service Time Estimation

The remaining service time is calculated for both non-edged and edged UEs with two different methods. In the first method, for non-edged UEs, the system will compute the range of the cell or coverage of the cell based on the RSRP of the A3 events λ_D. The serving small cells RSRPs (M_{A3} w) are computed from the A3 event measured report [5]. Therefore, the boundary of the cell is computed as follows:

$$\lambda_D = \frac{1}{k}\sum_{i=1}^{k} M_{A3_S} \quad (4)$$

Remaining service time is computed when UEs experience a decrease in received RSRP. Consequently, there will be ongoing monitoring of declining RSRP values from the serving cell. When the signal quality in the serving cell falls below a specific threshold, it can initiate an A2 event to relay the RSRP received by the UEs, enabling the monitoring of RSRP for non-edge UEs..

The estimated RT of a UE is based on the RSRP measurements gathered during the event A2. When the user environment UE is moving outwords, it reports that $\lambda_{p1} > \lambda_{p2}$, where λ_{p1} and λ_{p2} represents the measurement report of RSRP at time v_1 and v_2 . Therefore, the reduction rate of RSRP is computed over the time interval **Δ**v is given as follows:

$$\mu = \frac{1}{\Delta v}(\lambda_{p2} - \lambda_{p1}) \quad (7)$$

Therefore, the remaining service time of non-edged UEs are computed as

$$S_{neUE} = \frac{1}{\mu}(\lambda_{p2} - \lambda_D) \quad (8)$$

The system will adopt the A4 event to the edge-UE as the first threshold (λ_{A4}) which is related to the first measurement report λ_{p1} of RSRP A2 event. Along with that, a new measuring report λ_{A3A4} of RSRP is added with is corresponding to the second measurement report λ_{p2} of RSRP A2 event. Consequently, we can calculate the reduction rate of RSRP, and the remaining service time of the edge UE is computed as follows:

$$S_{eUE} = \frac{1}{\mu}(\lambda_{A3A4} - \lambda_D) \quad (9)$$

B. UE Session Time Estimation

The UE session time (Q) is subjected to eNB data rate, size of the application, maximum and minimum data rates. Therefore, it is calculated as:

$$Q = \frac{Y}{r} \quad (10)$$

Where Y represents the application data size in bits, r represents the allocated data rate in bits per second. The authors in [4] discussed the session time and the issues which affects it. The UE remaining session time is calculated based on the load imposed by the UE shortly.

$$R_Q = Q - C \quad (11)$$

Where C represents the consumed time where the load of the UE imposed previously. It is to remember that the eNB can compute the C at any time of the UE session.

C. Service Time at Target Cell (S_T)

The service time at the target cell is determined by estimating the duration the UE will stay in the neighbouring target cell. This estimation relies on the UE's average speed (G) and the coverage range of the small cell (Z).

$$S_T = Z / G \quad (12)$$

IV. PROPOSED LOAD BALANCING ALGORITHM

A. Network Monitoring for Data Gathering

The Network is monitored with c-SON and collects the information periodically. If a small cell exceeds its adaptive threshold, it should be considered overloaded. Then, the system will force some of the UEs to be handover to the neighbouring cells or Macro which are lightly loaded.

The c-SON gathers measurement reports from UEs in different serving cells. When cells experience heavy loads, they transfer candidate UEs to less burdened serving cells. Based on the data collected, the three sets of UEs are categorized into fast, slow, and very slow.

Let us consider that Ue-fast= {Ue-fast1, Ue-fast2, …}, Une-fast= {Une-fast1, Une-fast2, …} represents the fast set of edge and non-edge UEs, where Ufast= { Ue-fast, Une-fast}, Uslow= {Uslow1, Uslow2, …}, Uvery-slow= {Uvery-slow1, Uvery-slow2, …} represents the very slow set of UEs. Here, the candidate UEs are selected based on the fast and slow UEs UcandidateUEs={Ufast, Uslow}. The candidate UEs are organized in ascending order according to the RSRP values of their serving cells. Additionally, the c-SON generates another set that includes neighboring target cells.

B. Utility Based Load Balancing Algorithm for Hand Over Process

The c-SON runs the proposed algorithm periodically to perform the load balancing operation. Algorithm 1 explains about the utility based load balancing algorithm for reducing frequent handovers. The c-SON will perform handovers of candidate set UEs from overloaded cells to underloaded neighboring cells or the Macrocell based on considerations such as the UEs remaining session time and service time and utility function. In the initial stage, the small cells forward the load information of resource block utility function (LC) to the c-SON. In the next step, the small cells are arranged in the descending order based on the LC.

Algorithm 1: Propoosed Algorithm for Handover Process

Compute LCmacro and LCw
Identify the overloaded small cells and create the set M
Compute Ek for each cell in the set M and sort them in the descending order
Classify UEs based on the Eqs. 16,17,18 and 19
For all m∈M do
  Compute the maximum load µm of the cells using Eq. 20
    For all Uveryslow do
      No handover proves is initiated
    End for
    If LCmacro<Tmacro then
      For all Ufast do
        If LCmacro+µUfast < Tmacro then
      Handover is initiated for UEs to the macrocell
      End if
      End for
    For all Uslow do
      Perform load balancing process [2]
    End for
End for

Subsequently, the proposed algorithm evaluates the highest load in the list against a predefined threshold value. If the cell's load surpasses the threshold value, it signals an overload status to the serving cell, necessitating the initiation of the load balancing process. To make the algorithm more adaptive to the network, the threshold values are as adaptive threshold Tad which is given as follows:

$$T_{ad} = \max(\overline{LC_{nw}}, T_{in}) \tag{13}$$

For each cell, the current load LC and adaptive threshold Tad is compared to find the load status. The algorithm generates the new set M which contains the overloaded cells which satisfies the condition in Eq. 14.

$$\overline{LC_m} \geq T_{in} \quad for\ m \in M; M \subset w \tag{14}$$

The c-SON rearranges the set M in descending order and computes the efficiency factor using Eq. 15.

$$E_k = \sum_{k \subset B_k} \min(\sum_j \mu(k, j), (1 - \overline{LC_k})) \tag{15}$$

Then, the system will take each overloaded cell and reduces its load by transferring their candidate UEs to the lightly loaded cells or Macro. UEs within each overloaded small cell are categorized as fast, slow, or very slow. If the UE satisfies the Eq.16 then it is declared as the fast moving.

$$R_Q > S \tag{16}$$

However, the fast moving edge UEs must meet the following condition:

$$R_Q > (S + S_T) \tag{17}$$

The slow moving edge UEs must meet the following condition:

$$S < R_Q > (S + S_T) \tag{18}$$

Finally, the very slow moving edge UEs must meet the following condition:

$$R_Q < S \tag{19}$$

The algorithm computes the maximum load of each overloaded cell which has to be moved to the target neighbouring cell based on the following equation.

$$\tilde{\mu}_m = \overline{LC_{nw}} - \frac{1}{2}\mu(M, e_m^j) \quad (19)$$

The algorithm considers the very slow UE set and there will be no handover is performed. Because, their session should be completed before they exit the serving cell. This will avoid the number of handovers. In the next stage, the load of the macro is defined as LCmacro and the threshold for macro is fixed by the operator Tmacro. Then the algorithm will takes the fast UEs from the UE candidate set and forwards them to the Macrocell based on the Eq. 20.

$$LC_{macro} + \hat{\mu}_{U_{fast}} < T_{macro} \quad (20)$$

Where µUfast is the assessed load of the fast moving UE Ufast at the Macrocell. Simultaneously, the small cell handovers are avoided.

As per the MLB algorithm [2], the slow UEs are transferred lightly loaded or normal loaded small cells. Finally, the algorithm will update all the parameters related to the cell.

.

V. Experimental Analysis

The proposed algorithm's performance is evaluated through testing with the MATLAB simulator. A heterogeneous network is considered with 10 small cells, 70UEs and 1 Macrocell. UEs are divided in to two categories. In the first category, the UEs utilize the Internet for browsing, and the data size for each session is specified as 3MB and the data rate of 128 kbps as minimum value, 256 kbps as mean value and 512 kbps as maximum value. In the second category, the UEs are using the 4K video streaming with data size of 2.5GB and the requested data rate is 512 Mbps as minimum value, 1Gbps ad mean value and 4Gbps as maximum value. Table 1 shows the video streaming parameters. The bandwidth allocated for each cell is 20 MHz and the available resources for each cell are 100 PRBs. The transmission power small cell is given as 24dBm and for Macrocell is 46dBm. Table 2 shows the parameter settings for the network.

Table 1: Video Streaming Parameters

Parameters	Value
Bitrate Adaptation	512 kbps
Buffer Size	5 Sec
Latency	50 ms
Frame rate	30 fps
Pre-fetching	10 sec
Forward Error Correction	10%
Length of the Video	1 hour
Type of the Video	Live TV show
Stored Video	Movie clip

Table 2: Parameter Settings for Heterogeneous Network

Parameter	Value
Small Cell Transmission Power	24 dBm
Macro Cell Transmission Power	46 dBm
Bandwidth allocation for cell	20 MHz
Background Load of Macro Cell	0.2
Small Cells	10
UEs	70
Macro Cell	1
Scheduler	CQAff

Fading Channel	Log-normal, 4dB standard deviation
Hysteresis	2dB
Threshold for small Cell (Tint)	0.75
Threshold for Macro Cell (Tmacro)	0.8

In the proposed model, we consider that the data size is fixed at the given time period per UE in the downlink. Two type of UEs mobility is considered such as urban mobility (Pedestrian Speeds) with a speed of 3.5 km/h and along with that suburban mobility (vehicle speeds) with a speed of 30 km/hr. The UEs are distributed randomly in the small cells.

Figure 2 shows the performance comparison of different load balancing algorithms with respect to standard deviation of the Load across the network. was noted that the proposed algorithm exhibited a lower standard deviation in comparison to the other algorithms. The proposed algorithm recorded the 81.4%, 78.2% and 34.3% standard deviation reduction compared to NoMLB, FMLB [14] and DMLB [21] respectively. Therefore, the small cells load variance is minimized and system is well balanced.

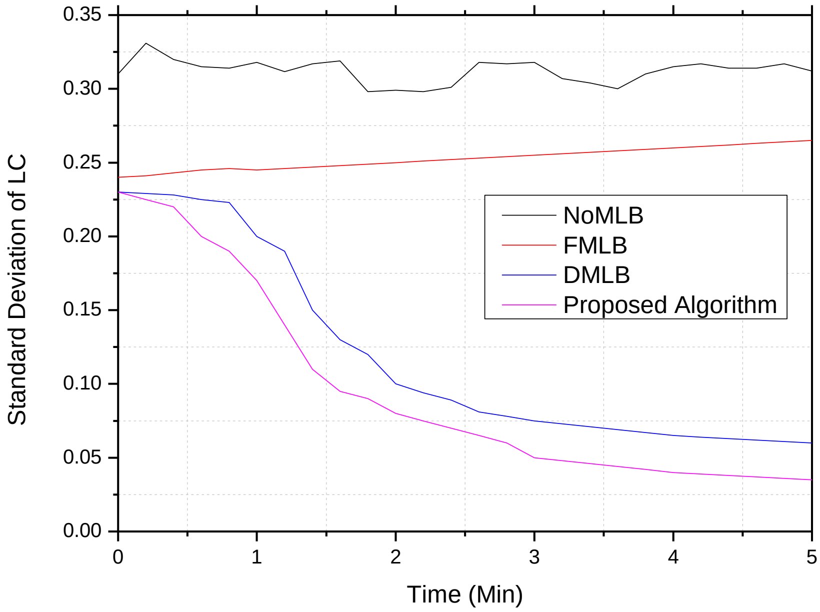

Figure 2: Standard Deviation of the of the Load in the Small Cells

Figure 3 shows the average data rate of UE for different mobility load balancing algorithms. The proposed algorithm achieved a 51% success rate for UEs with an average data rate exceeding 1Gbps. In contrast, the DMLB algorithm achieved a 35% success rate, the FMLB algorithm achieved 1.15%, and the NoMLB algorithm achieved 0.5% for UEs with an average data rate exceeding 1Gbps.

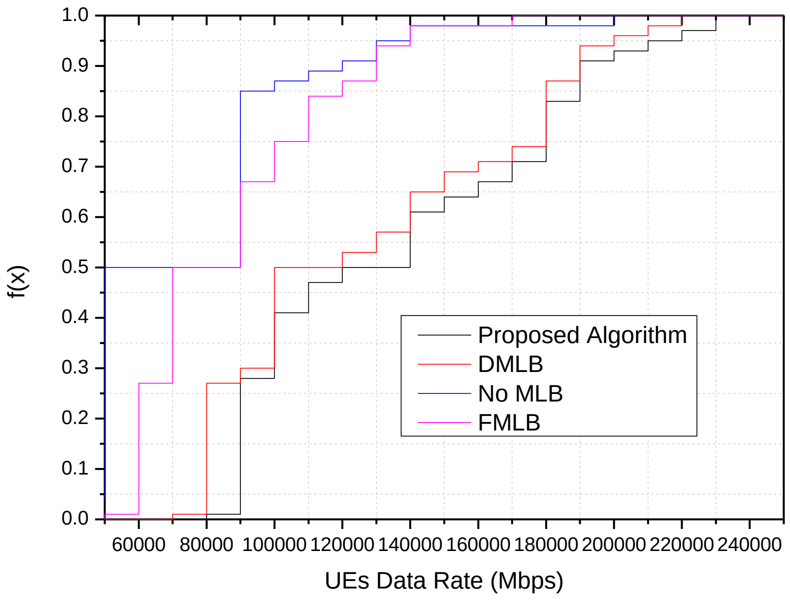

Figure 3: Average Data Rate of User Environments

The count of the handovers that the MLB algorithms performed during a given period is evaluated in Figure 4. It is obvious that the DMLB doesn't categorize the various types of UEs, but it takes into account their performance evaluation. We can detect the types of these through the simulation. Based on the procedure specified by the DMLB, we can observe the various types of handovers that the different UEs performed. We then compare the proposed algorithm with the procedure specified by the DMLB. Figure 4 reveals that the proposed algorithm effectively reduces both forced and normal handovers by 64.7% and 39.8%, for FMLB and DMLB respectively.

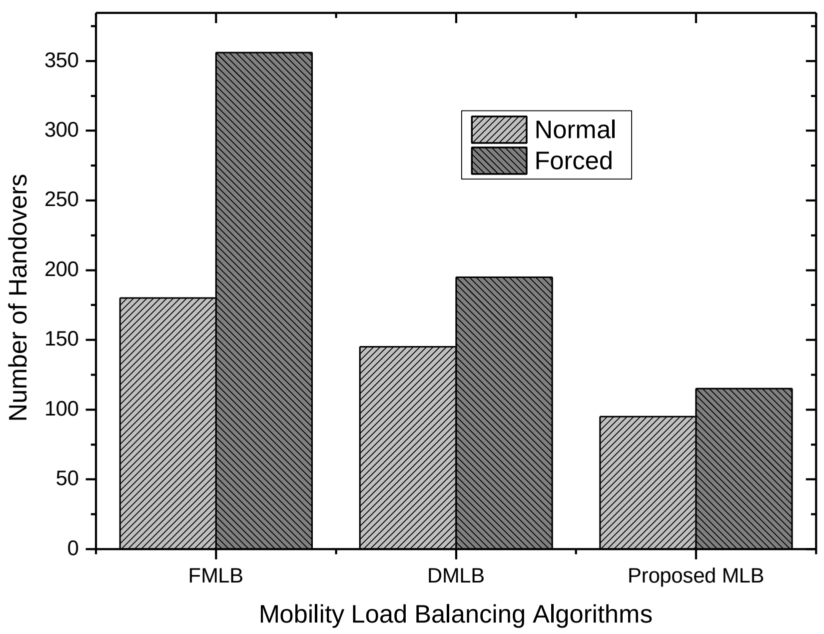

Figure 4: Mobility Load Balancing Algorithms Vs Handovers

.

VI. Conclusion

This paper proposed load balancing algorithm for resource management with minimum handover process for 5G networks. This method takes into account not only the edge UEs but also the non-edge UEs for load balancing within small cells. The major objective the proposed model is to transfer the overloaded small cells load to the lightly loaded macro cells or normal cell by hand over the fast-moving UEs. n the proposed method, UEs are categorized as fast, slow, or very slow movers and then transferred to lightly loaded macro or small cells, effectively preventing unnecessary handover processes. The algorithm considers the different types of edges of the network and the non-edges of the network for handling overloaded small cells. If the UE is moving fast, the fast-moving UE will be transferred to a macro cell, while the slow-moving one will be handed over to a neighboring small cell. If the UE is moving very slowly, it will not be transferred to the other cell. This ensures that the unnecessary handover doesn't occur. The proposed algorithm achieves the smallest possible number of handovers to ensure that the load balancing process is performed efficiently.

REFERENCES

[1] Saad, Wasan Kadhim, Ibraheem Shayea, Abdulraqeb Alhammadi, Muntasir Mohammad Sheikh, and Ayman A. El-Saleh. "Handover and load balancing self-optimization models in 5G mobile networks." Engineering Science and Technology, an International Journal 42 (2023): 101418.

[2] K. M. Addali, S. Y. B. Melhem, Y. Khamayseh, Z. Zhang, and M. Kadoch, "Dynamic Mobility Load Balancing for 5G Small-Cell Networks Based on Utility Functions," IEEE Access, vol. 7, pp. 126998-127011, 2019.

[3] Riyazuddien, Shaik, Dhulipalla Venkata Rao, Madhu Ramarakula, Satamraju Krishna Prasad, and Krishna Chennakesava Rao Madaka. "A SELF-TUNING ALGORITHM FOR OPTIMAL ENERGY EFFICIENT LOAD BALANCING IN WIRELESS CELLULAR HETEROGENEOUS NETWORKS." Telecommunications and Radio Engineering 82, no. 1 (2023).

[4] S. K. Albasheir, "Enhanced resource utilization in the evolved packet core network for 4G/LTE mobile broadband services," École de technologie supérieure, 2016.

[5] M. M. Hasan, S. Kwon, and S. Oh, "Frequent handover mitigation in ultra-dense heterogeneous networks," IEEE Transactions on Vehicular Technology, vol. 68, no. 1, pp. 1035-1040, 2018.

[6] Wu, Chia-Lun, Tsung-Tao Lu, Chin-Tan Lee, Jwo-Shiun Sun, Hsin-Piao Lin, Yuh-Shyan Hwang, and Wen-Tsai Sung. "Evolution towards Coordinated Multi-Point Architecture in Self-Organizing Networks for Small Cell Enhancement Systems." Electronics 12, no. 11 (2023): 2473.

[7] Y. Q. Bian and D. Rao, ``Small cells big opportunities," Huawei Technol., Shenzhen, China, Tech. Rep., Feb. 2014.

[8] S. Feng and E. Seidel, ``Self-organizing networks (SON) in 3GPP long term evolution," Nomor Res. GmbH, Munich, Germany, Tech. Rep., May 2008.

[9] R. Fedrizzi, L. Goratti, T. Rasheed, and S. Kandeepan, ``A heuristic approach to mobility robustness in 4G LTE public safety networks," in Proc. IEEE Wireless Commun. Netw. Conf., Apr. 2016, pp. 1-6.

[10] R. Kwan, R. Arnott, R. Paterson, R. Trivisonno, and M. Kubota, ``On mobility load balancing for LTE systems," in Proc. IEEE 72nd Veh. Technol. Conf.-Fall, Sep. 2010, pp. 1-5.

[11] Hasan, Md Mehedi, Sungoh Kwon, and Jee-Hyeon Na. "Adaptive mobility load balancing algorithm for LTE small-cell networks." *IEEE transactions on wireless communications* 17, no. 4 (2018): 2205-2217.

[12] Ma, Bo, Bowei Yang, Yunpeng Zhu, and Jie Zhang. "Context-aware proactive 5G load balancing and optimization for urban areas." *IEEE Access* 8 (2020): 8405-8417.

[13] Xu, Y., Xu, W., Wang, Z., Lin, J. and Cui, S., 2019. Load balancing for ultradense networks: A deep reinforcement learning-based approach. *IEEE Internet of Things Journal*, *6*(6), pp.9399-9412.

[14] Zhang, Hongliang, Lingyang Song, and Ying Jun Zhang. "Load balancing for 5G ultra-dense networks using device-to-device communications." *IEEE Transactions on Wireless Communications* 17, no. 6 (2018): 4039-4050.

[15] Vu, Trung Kien, Mehdi Bennis, Sumudu Samarakoon, Mérouane Debbah, and Matti Latva-Aho. "Joint load balancing and interference mitigation in 5G heterogeneous networks." *IEEE Transactions on Wireless Communications* 16, no. 9 (2017): 6032-6046.

[16] Wang, Tianyu, Shaowei Wang, and Zhi-Hua Zhou. "Machine learning for 5G and beyond: From model-based to data-driven mobile wireless networks." *China Communications* 16, no. 1 (2019): 165-175.

[17] Nguyen, Van-Giang, Karl-Johan Grinnemo, Javid Taheri, and Anna Brunstrom. "On load balancing for a virtual and distributed MME in the 5G core." In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1-7. IEEE, 2018.

[18] Shayea, Ibraheem, Mustafa Ergen, Marwan Hadri Azmi, Sultan Aldirmaz Çolak, Rosdiadee Nordin, and Yousef Ibrahim Daradkeh. "Key challenges, drivers and solutions for mobility management in 5G networks: A survey." *IEEE Access* 8 (2020): 172534-172552.

[19] M. M. Hasan, S. Kwon, and J.-H. Na, ``Adaptive mobility load balancing algorithm for LTE small-cell networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2205-2217, Apr. 2018.

[20] S. Oh, H. Kim, and Y. Kim, ``User mobility impacts to mobility load balancing for self-organizing network over LTE system," in Proc. 14th Int. Conf. Adv. Trends Radioelectron., Telecommun. Comput. Eng. (TCSET), Feb. 2018, pp. 1082-1086.

[21] Sun, Songlin, Liang Gong, Bo Rong, and Kejie Lu. "An intelligent SDN framework for 5G heterogeneous networks." *IEEE Communications Magazine* 53, no. 11 (2015): 142-147.