QoS Aware User Association in Massive MIMO Enabled Hetnets for DTU and NDTU Traffic

Md. Kamran Chowdhury Shisher[1], Tasmeen Zaman Ornee[2] and Md. Farhad Hossain[3]
Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
Email: 1541kamran@gmail.com[1], tzornee@gmail.com[2] and mfarhadhossain@eee.buet.ac.bd[3]

***Abstract*—Massive MIMO and heterogeneous networks are the two promising candidates for the future fifth generation (5G) cellular systems. User association policy is one of the most vital issues for the performance of this network. Therefore, this paper proposes a novel user association technique for massive MIMO based heterogeneous cellular networks with co-existing delay tolerant user (DTU) and non-delay tolerant user (NDTU). Proposed technique aims to improve the quality of service (QoS) by accounting the signal quality, call blocking probability and waiting time for associating the user equipment (UE) with appropriate base stations (BSs). The system is modeled using M/M/c/N queue. System performance is evaluated in terms of performance metrics, namely, throughput, call blocking probability and waiting time. Impact of system parameters, such as UE arrival rate, tolerable waiting time and number of antenna are also investigated.**

***Index Terms*—User association, massive MIMO, heterogeneous network, DTU and NDTU traffic, QoS.**

I. INTRODUCTION

Future wireless cellular networks will face thousand times growth of mobile data traffic. A recent study investigated that 25 billion interconnected devices are anticipated within 2020 and 7 billion cumulative smartphones are projected between 2013-2017 [1]. At present and in the near future, to meet this demand, we need more spectrum, more efficiency and better quality of service (QoS). In light of this, several approaches have drawn considerable attention. Few of them are, massive MIMO, heterogeneous network (HetNet), mmWave network, energy harvesting network, device to device communication (D2D), cloud radio access network (C-RAN) and full duplex communication [2]–[6]. The focus of this paper is massive MIMO enabled heterogeneous cellular network. Massive MIMO is a system where number of antenna is significantly greater than number of user [3], [7]. This can support high spectral efficiency with simple linear transceivers and is expected to provide high energy efficiency [6]. On the other hand, heterogeneous network (HetNet) is to deploy small cell access points (SCAs) in the coverage of a Macro BS [3]. Spatial reuse and coverage can be improved significantly by deploying additional network nodes within the local area range and bringing the network closer to end users [6].

For achieving the best performance from such networks, efficient user association policy is required [2]. Therefore, this area has drawn a considerable attention of the research community [2], [8]–[13]. For user association in 5G networks, different metrics have been adopted. Five common metrics are: outage/coverage probability, spectral efficiency, energy efficiency, quality of service (QoS) and fairness [2]. QoS can be measured in terms of blocking probability, traffic delay and throughput. Many researchers are looking forward to develop an effective user association for future network. In [8], [9], user association policy is adopted for heterogeneous network. In paper [8], the authors designed user association scheme for downlink heterogeneous network and opted to improve energy efficiency. Whereas, the main goal of user association in [9] is the maximization of the sum data rate of all users, under users minimal rate constraint considering user fairness. This paper adopted Hungarian algorithm to implement their user association policy. On the other hand, authors in [10] designed a resource efficient load balancing user association in massive MIMO. They worked with two user centric schemes, namely, max-rate association and load based association. In max-rate association, user decides to associate in peak rate but in load based scheme, user tries to selfishly associate to maximize throughput by also considering load information. In [11]–[13], user association scheme was investigated for massive MIMO enabled HetNets. The objective of user association in [11] was to maximize system capacity. They also considered limited load capacity at each BS, without allowing fractional user association. In [12], authors adopted a QoS aware user association in a game theoretic way. Their main QoS parameter was data rate. In addition to the improvement of data rate, they also worked for antenna allocation for different class of users. Moreover, authors in [13] proposed user association policy to achieve energy efficiency for massive MIMO and mmWave enabled heterogeneous networks. To acquire energy efficient network, they also considered BSs with the capability of harvesting renewable energy.

The increasing sum data rate in massive MIMO involves increasing number of antenna. But the increased number of antennas consume more power and need more radio resources. In light of this issue, in [14], author proposed to restrict antenna to user ratio below ten. The limited antenna and radio resources also limit the number of user served at a time. In the busy period, some users may have to be blocked from getting service which affects QoS of the system. Moreover, in practice there is a mix scenario where both DTU and NDTU traffic co-exist. DTUs can delay before the starting of their service. But the long waiting time to get the service affects the QoS. Therefore, any efficient user association policy should consider both DTU and NDTU traffic. To the best of our knowledge, there is no such work of user association in massive MIMO enabled HetNets which considers both DTU and NDTU traffic. Furthermore, for future implementation, a perfect user association policy needs to consider multiple QoS parameters. But the papers discussed above considered only throughput.

In light of these, this paper proposes a novel user association mechanism considering both DTU and NDTU traffic. The proposed policy jointly considers multiple QoS parameters, namely, blocking probability, waiting time and throughput. Our simulation results show throughput, waiting time and blocking probability with varying arrival rate and different traffic scenarios. Moreover, impact of system parameters is also demonstrated through simulation of our algorithm. This paper also compares performance of the proposed user association scheme with that of only SINR based scheme demonstrating the superiority of the proposed algorithm.

The remainder of this paper is organized as follows: Section II presents system model, section III formulates the problem and describes the algorithm, in section IV numerical results and analysis are given and section V concludes our work summarizing the key findings.

II. System Model

A. Network Model

In this paper, we consider the downlink of a Massive MIMO enabled heterogeneous cellular network. This system consists of one macro BS and several micro BS randomly distributed inside it. All BSs are equipped with large number of antennas for enabling massive MIMO communication. The antennas in the BS are employed linearly having no correlation and mutual coupling between transmitter and receiver antennas [15]. User equipment (UE) are assumed to be uniformly distributed throughout the network area. Two types of UE are considered, namely, delay tolerant user (DTU) and non-delay tolerant user (NDTU). DTUs accept to delay their service in case all the servers are busy. NDTUs are not willing to wait under any circumstances. Transmission of users from all BSs are served at the same time occupying the same bandwidth (BW). We assume perfect CSI acquisition. Network model of the system is illustrated in Fig. 1.

B. Link Model

In this paper, we consider Rayleigh channel model [15]. The $K \times M$ channel matrix of j^{th} BS can be modeled as

$$H_j = G_j^{\frac{1}{2}} I_j \tag{1}$$

where G_j is $diag[\beta_{j,1}, \beta_{j,2}, ...\beta_{j,k}]$ where $\beta_{j,k}$ and I_j $(K \times M)$ matrix can be expressed as

$$\beta_{j,k} = g_{j,t} g_{j,r} d_{j,k}^{-\alpha} f_{j,k} \tag{2}$$

$$I_j = [I_{j,1}, I_{j,2}, ..., I_{j,K}] \tag{3}$$

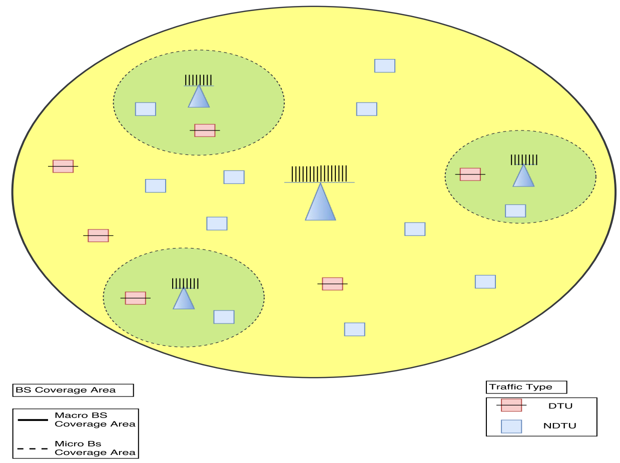

Fig. 1. Network model

where $I_{j,k}$ vector can be expressed as

$$I_{j,k} = [I_{j,k,1}, I_{j,k,2}, ..., I_{j,k,M}] \tag{4}$$

here $g_{j,t}$ and $g_{j,r}$ are transmitter and receiver antenna gain respectively, $d_{j,k}$ is the distance between BS j and user k and α is the path loss component, $f_{j,k}$ and $I_{j,k,n}$ models the shadow fading and Rayleigh fading respectively. K is the total user and M is the total number of antenna. Channel matrix of the j^{th} BS can be given as

$$H_j = [H_{j,1}, H_{j,2}, ..., H_{j,k}] \tag{5}$$

Signal-to-interference-plus-noise-ratio (SINR) from BS j to user k can be expressed as

$$SINR_{j,k} = \frac{P_{rec,j,k}}{I_{inter,k} + I_{intra,k} + P_n} \tag{6}$$

where $P_{rec,j,k}$ is the received signal power to user k from j^{th} BS, $I_{inter,k}$ is the interchannel interference, $I_{intra,k}$ is the intrachannel interference and P_n is the noise power.

C. Proposed User Association Algorithm

In our algorithm, we propose sequential consideration of blocking probability, waiting time and SINR respectively. To maintain communication, a threshold SINR, S_t is taken. BSs that provide SINR greater than or equal to S_t are considered primary candidate BSs for associating a UE. In our proposed scheme, firstly, the primary candidate BSs are selected. Secondly, server of the primary candidate BSs are checked. Among the primary candidate BSs which have free server are selected as secondary candidate BSs for associating a UE. At last, the BS which provides maximum SINR among the secondary candidate BSs will be selected for associating a UE. If every server of all BSs is found busy, NDTU will be blocked. However, DTU will not be blocked as they can be assigned in the queue of a BS.

In case of a DTU, when all server is found busy, waiting time of the DTU in every BS is calculated. Among the primary

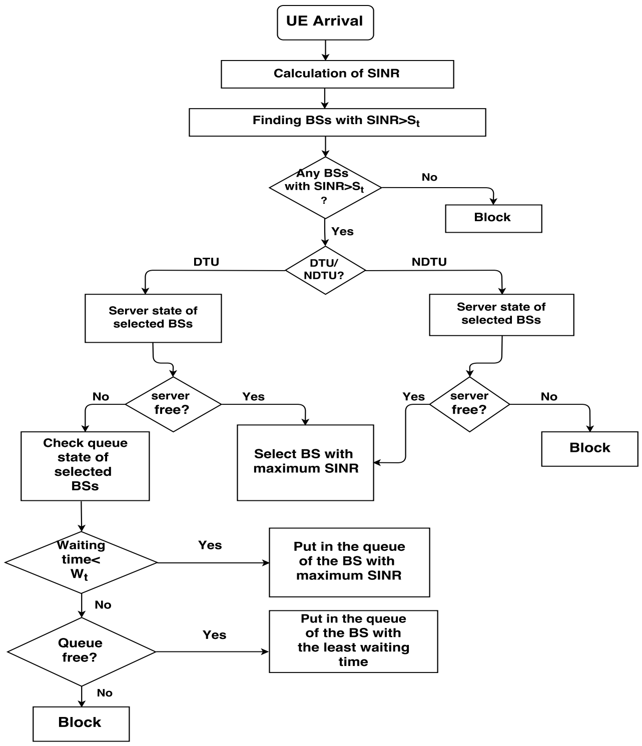

Fig. 2. Flow chart of the proposed user association algorithm

candidate BSs, those provides waiting time less than W_t are selected for secondary candidate BSs. Then BS which provides maximum SINR among the secondary candidate BSs is assigned for the DTU.

Furthermore, if all server of every BS are busy and every BS provides waiting time greater than W_t, there will be no secondary candidate BS. In this case, the queue of all BSs is checked. After that, among the primary candidate BSs which provides minimum waiting time with available queue will be assigned to DTU. At last, if there is no available queue, no available server, DTU will be blocked.

Our algorithm is illustrated in the flow chart in Fig. 2. Pseudocode of the proposed algorithm is also shown in Algorithm 1. In Algortihm. 1 PC is denoted as primary candidate BS, SC is secondary candidate BS, Q is the number of UE in queue, S is the number of busy server and Q_t is the threshold number of UE in queue.

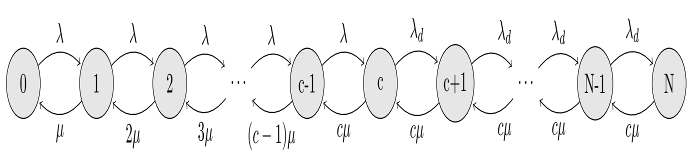

Fig. 3. Markov chain for each BS

Algorithm 1 Pseudocode of the proposed algorithm

initialize PC, SC, S_t, Q_t, S and Q
if k^{th} UE requests service **then**
  for $j = 1 : J$ **do**
    Compute SINR of k^{th} user
  end for
  for $j = 1 : J$ **do**
    if $SINR_{j,k} \geq S_t$ **then**
      Select j^{th} BS as PC
    end if
  end for
  Compute S of PC BSs
  for All PC BSs **do**
    if The BS has free server **then**
      Select the BSs as SC
    end if
  end for
  if There is atleast one SC BS **then**
    Associate k^{th} UE to BS with max SINR among SC
  else
    if k^{th} UE is NDTU **then**
      k^{th} UE will be blocked
    else
      Compute Q of PC BSs
      for ALL PC BSs **do**
        if $Q \leq Q_t$ **then**
          Select j^{th} BS as SC
        end if
      end for
      if There is atleast one SC BS **then**
        Associate k^{th} UE to BS with max SINR among SC
      else if There is no free queue in any BS **then**
        DTU will be blocked
      else
        Associate k^{th} UE to BS with min Q among PC
      end if
    end if
  end if
  Update all S, Q, PC, SC according to BS selected
end if

D. Queue Model

To model the proposed algorithm, we adopt queue theory based approach. Here, M/M/c/N queue model of system length N with c number of server and $N - c$ queue length has been applied. Queue theory has been adopted for DTU traffic and for NDTU traffic, N is equal to c and there is no queueing option for this case. The method of service is First Input First Output (FIFO) with considering priority of NDTU over DTU. Markov Chain for the queue model for each BS is shown in Fig. 3.

For the queue model, following three situations can arise.

Situation 1 (Available server, available queue): When

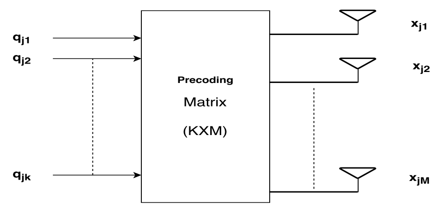

Fig. 4. Massive MIMO transmitter configuration

a server becomes free and DTUs are waiting in the queue, according to FIFO method, in this situation, UE which has come first among the DTUs waiting, will get the srrvice first. But at this time, if a NDTU comes, as NDTU has more priority over DTU, it will get service before any user waiting in the queue.

Situation 2 (Full server, available queue): In this case, DTUs will enter the queue and wait until any server gets free. On the other hand, NDTUs will be blocked instantly.

Situation 3 (Full server, full queue): In this case, any user coming will be blocked.

III. Performance Metrics

This section defines the performance metrics used in this paper.

A. Blocking Probability

Blocking probability is given the highest priority among all the metrics. For achieving better performance of any system, it is necessary to decrease the blocking probability as much as possible. Blocking probability P_b can be given as

$$P_b = \frac{K_b}{K} \tag{7}$$

where K_b is the number of user blocked and K is the total number of arrived user.

B. Waiting Time

Waiting time is another important QoS parameter for DTUs. Though DTU can wait to get the service, it cannot afford long waiting time. More waiting time for DTU in queue deteriorates the system overall QoS. In light of the fact, this paper considers threshold waiting time, W_t. Average waiting time W_q and threshold waiting time W_t can be expressed as

$$W_q = \frac{L_q}{\lambda_d} \tag{8}$$

$$W_t = \frac{Q_t}{\lambda_d} \tag{9}$$

where L_q is the average queue length, Q_t is the threshold number of user in the queue and λ_d is the arrival rate of DTU.

C. Throughput

Transmitter configuration with massive MIMO is shown in Fig. 4 where $\mathbf{q}_j$ is the vector of input data stream from BS j given by

$$\mathbf{q}_j = [q_{j1}, q_{j2}, ..., q_{jK}] \tag{10}$$

Then q_{jk} is input data from BS j of user k.

This data stream passes through the precoding matrix and produces transmit antenna output vector of j^{th} BS $\mathbf{X_j}$. $\mathbf{X_j}$ can be expressed as

$$\mathbf{X_j} = \mathbf{q}_j \mathsf{W_j}^T \tag{11}$$

where $\mathsf{W_j}$ $(K \times M)$ is the precoding matrix of BS j and $\mathbf{W}_{jk}$ is the precoding vector of BS j for user k.

Signal for user k from BS j can be expressed as

$$Y_{j,k} = \sqrt{\alpha P_j} \mathbf{H}_{j,k} \mathbf{X} + I_{inter,k} + n_{j,k} \tag{12}$$

where k is 1, 2, 3, ..., K active users, $n_{j,k}$ is the thermal noise and α is a normalization constant [4].

Now, $Y_{j,k}$ can then be written as

$$Y_{j,k} = \sqrt{\alpha P_j} \mathbf{H}_{j,k} \mathbf{W}_{j,k} \mathbf{q}_{j,k} + \sqrt{\alpha P_j} \sum_{k' \neq k}^{K_j} \mathbf{H}_{j,k} \mathbf{W}_{j,k'} \mathbf{q}_{j,k'}$$
$$+ \sum_{j' \neq j}^{J} \sum_{k' \neq k}^{K_j'} \sqrt{\alpha P_{j'}} \mathbf{H}_{j'k} \mathbf{W}_{j'k'} \mathbf{q}_{j',k'} + n_{j,k} \tag{13}$$

Here, Interference experienced by user k are

$$I_{inter,k} = \sum_{j' \neq j}^{J} \sum_{k' \neq k}^{K_j'} \sqrt{\alpha P_{j'}} \mathbf{H}_{j'k} \mathbf{W}_{j'k'} \mathbf{q}_{j',k'} \tag{14}$$

$$I_{intra,k} = \sqrt{\alpha P_j} \sum_{k' \neq k}^{K_j} \mathbf{H}_{j,k} \mathbf{W}_{j,k'} \mathbf{q}_{j,k'} \tag{15}$$

Therefore, SINR of k^{th} user from j^{th} BS can then be expressed as

$SINR_{j,k} =$

$$\frac{\alpha P_j |\mathbf{H}_{j,k} \mathbf{W}_{j,k}^T|^2}{\alpha P_j \sum_{k' \neq k}^{K_j} |\mathbf{H}_{j,k} \mathbf{W}_{j,k'}|^2 + \sum_{j' \neq j}^{J} \sum_{k' \neq k}^{K_{j'}} \alpha P_{j'} |\mathbf{H}_{j',k} \mathbf{W}_{j',k'}|^2 + n_{j,k}} \tag{16}$$

where precoding matrix is used according to zero force beamforming (ZFBF) precoding code [4], [16] and can be expressed as

$$\mathsf{W_j} = \mathsf{H_j}^* (\mathsf{H_j}^T \mathsf{H_j}^*)^{-1} \tag{17}$$

It has following property [15], [16]

$$\frac{1}{M}H_{j,k}W_{j,k}^{T} = \beta_{j,k} \tag{18}$$

$$\frac{1}{M}H_{j,k}W_{j,k'}^{T} = 0 \tag{19}$$

$$\frac{1}{M}H_{j,k}W_{j',k'}^{T} = 0 \tag{20}$$

Then downlink system capacity can be written as

$$C = \sum_{k=1}^{K} \log_2(1 + SINR_k) \tag{21}$$

and with perfect CSI, ZFBF precoding matrix downlink system capacity can also be expressed as

$$C = \sum_{k=1}^{K} \log_2(1 + M\frac{P_j \beta_k}{n_{j,k}}) \tag{22}$$

IV. Numerical Result And Analysis

A. Simulation Setup

In this section, performance of the proposed algorithm through MATLAB based simulation is presented. The system is simulated with one macro BS and two micro BSs. Simulation is conducted for 2000 calls and the UE positions are generated uniformly distributed in space. Inter-arrival time and call duration are modeled with exponential distribution. Mean inter- arrival time is taken inverse of mean arrival rate. Mean call duration is taken 60 seconds. Number of NDTU is considered one third of total UEs. Antenna number to total server ratio is considered eight and Q_t is taken five unless otherwise specified. Moreover, results comparing proposed user association scheme with that of only SINR based policy are presented. The results presented in this section are the average of 20 iterations. Other simulation parameters are given in Table I.

TABLE I
Simulation Parameter

parameter	value
Bandwidth of Macro BS (MHz)	5
Bandwidth of Micro BS (MHz)	1.4
Transmit Power of Macro BS (dBm)	46
Transmit Power of Micro BS (dBm)	30
Standard Deviation of Shadowing Effect (dB)	8
Cell Radius of Macro BS (m)	500
Cell Radius of Micro BS (m)	50
Thermal Noise Power	kTB
Temperature (Kelvin)	300
Path loss exponent	3.7
Transmitter Antenna Gain	1
Receiver Antenna Gain	1
Call Duration ($\frac{1}{\mu}$) (seconds)	60
Macro BS Server	25
Micro BS Server	10
Macro BS buffer length	10
Micro BS buffer length	10

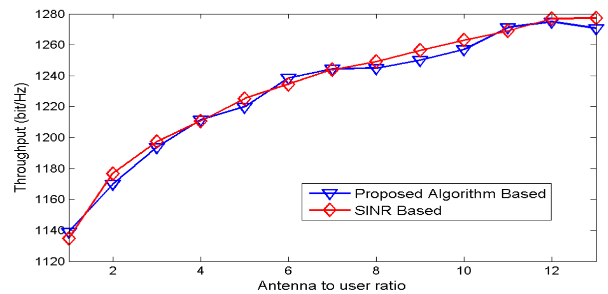

Fig. 5. Throughput of the system with different antenna to user ratio

B. Throughput

The impact of increasing number of antenna to user ratio on throughput of the system is presented in Fig. 5. This part of the simulation is conducted with the arrival rate of 0.7 user per second. With the increase of ratio, throughput increases, but the increasing trend is lower with the increase of antenna number. This is because, throughput is logarithmic with antenna number as seen in (22). This figure also demonstrates the comparison of SINR based scheme with the proposed scheme on throughput. It is evident that the throughput of the two schemes are very close to each other.

C. Blocking Probability

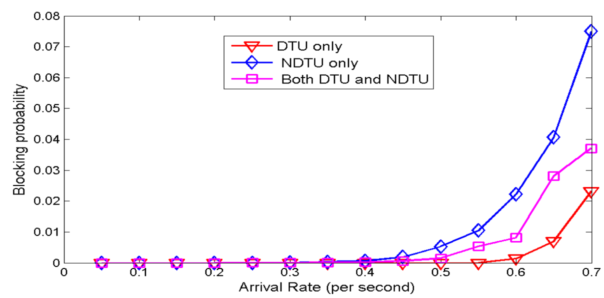

Fig. 6. Blocking Probability comparison for different types of traffic

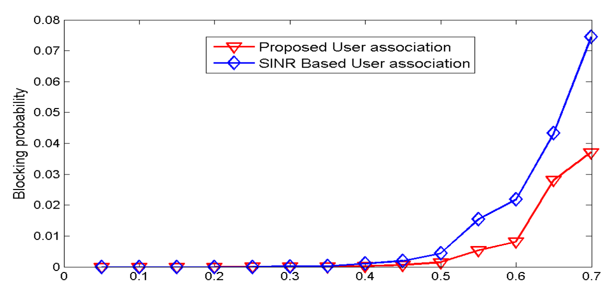

Fig. 7. Blocking probability comparison with proposed user association and SINR based association

In Fig. 6, call blocking probability under the proposed user association scheme is presented. This is investigated for three different traffic scenarios, namely, DTU, NDTU and mix scenerio of DTU and NDTU. With the increase of arrival rate, the blocking probability increases. It can be seen that the rate of increase in call blocking is the highest for NDTU only and the lowest for DTU only. But in real life, DTU and NDTU traffic co-exist.

We also compare blocking probability of SINR based user association scheme with the proposed scheme. Fig. 7 presents the comparison. The blocking probability is significantly lower in case of the proposed user association scheme. Our proposed scheme deploys queue to hold the DTU before getting service. Furthermore, even in case all the servers are busy and waiting time is more than threshold, our proposed scheme does not block any UE till the queue becomes full. For this reasons, our proposed scheme improves blocking probability.

D. Waiting Time

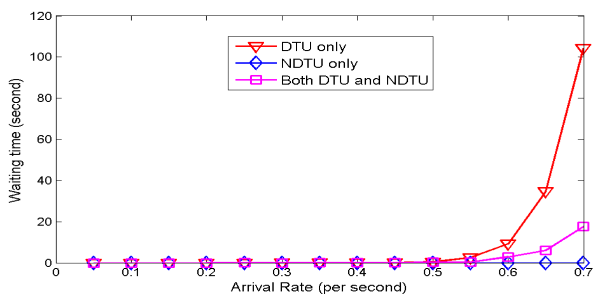

Fig. 8. Waiting time comparison for different types of user

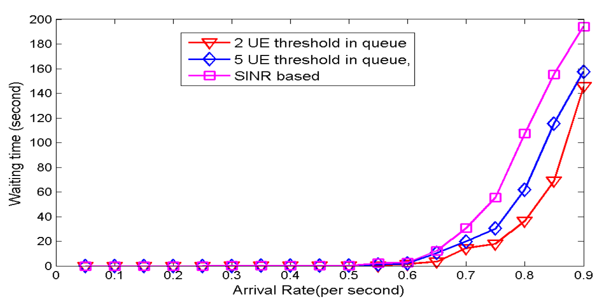

Fig. 9. Comparison of waiting time of proposed scheme with SINR based scheme

Fig. 8 shows the waiting time for the three scenarios with arrival rate. As NDTU does not wait in the queue, the waiting time for NDTUs only scenario is zero. For other two scenarios, waiting time increases with the increase of arrival rate. The increasing trend is more for the DTU only scenario.

Fig. 9 demonstrates the comparison of waiting time of the proposed policy with that of only SINR based. In SINR based policy, a UE is assigned in the BS which provides the best SINR. This policy may lead the UE to get the BS where it has to wait long time to get the service. This deteriorates QoS of the system. In our proposed policy, a UE is assigned to the BS where it needs to wait less, while gets a fair SINR. For this reason, this figure shows that waiting time in our proposed user association is better than that of SINR based. This figure also shows that with the decrease of Q_t, waiting time decreases.

V. Conclusion

In this paper, we have proposed a novel user association scheme for massive MIMO enabled heterogeneous cellular network. We have focused on the traffic scenario where both DTU and NDTU co-exist. Blocking probability, waiting time for DTU and throughput are critical QoS parameters of a cellular communication system. Conventional SINR based user association scheme considers only throughput. But in our proposed scheme, other two QoS parameters also have been taken into consideration. Simulation results demonstrate the improved blocking probability and waiting time without compromising throughput.

References

[1] Qualcomm, "The 1000x data challenge" Qualcomm Technol., Inc, San Diego, CA, USA, Tech. Rep, 2014. [Online]. Available: www.qualcomm.com/invention/technologies/1000x .

[2] Dantong Liu, Lifeng Wang, Yue Chen, Maged Elkashlan, Kai-Kit Wong, Robert Schober, Lajos Hanzo "User Association in 5G Networks: A Survey and an Outlook" *IEEE Communications Surveys and Tutorials* vol. 18, January, 2014, pp. 1018-1024.

[3] Emil Bjrnson, Marios Kountouris, Merouane Debbah "Massive MIMO and Small Cells: Improving Energy Efficiency by Optimal Soft-cell Coordination" *ICT 2013* October, 2013.

[4] Hien Quoc Ngo "Massive MIMO: Fundamentals and System Designs" 2015.

[5] Emil Bjrnson, Erik G. Larsson, Thomas L. Marzetta "Massive MIMO: Ten Myths and One Critical Question" *IEEE Communications Magazine* vol. 54, February, 2016, pp. 114-123.

[6] Ekram Hossain, Monowar Hasan "5G Cellular: Key Enabling Technologies and Research Challenges" *IEEE Instrumentation and Measurement Magazine* vol. 18, May, 2015, pp. 11-21.

[7] J. Hoydis, S. ten Brink and M. Debbah "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?" *IEEE J. Sel. Areas Commun.* vol. 31, no. 2, pp. 160171, 2013.

[8] Tianqing Zhou, Yongming Huang, Luxi Yang "Energy-efficient User Association in Downlink Heterogeneous Cellular Networks" *IET Communications* vol. 10, pp. 1553-1561, September, 2016.

[9] Dantong Liu, Yue Chen, Kok Keong Chai, Tiankui Zhang "Performance Evaluation of Nash Bargaining Solution Based User Association in HetNet" *International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)* pp. 571-577, 2013.

[10] Dilip Bethanabhotla, Ozgun Y. Bursalioglu, Haralabos C. Papadopoulos and Giuseppe Caire "User Association and Load Balancing for Cellular Massive MIMO" *Information Theory and Applications Workshop (ITA)* April, 2014.

[11] Yi Xu, Shiwen Mao "User Association in Massive MIMO HetNets" *IEEE Systems Journal* vol. 11, pp. 7-19, September, 2011.

[12] Ping Wang, Wei Song, Dusit Niyato, Yong Xiao "QoS-aware Cell Association in 5G Heterogeneous Networks with Massive MIMO" *IEEE Network* vol. 19, pp. 76-82, December, 2015.

[13] Bingyu Xu, Yue Chen, Maged Elkashlan, Tiankui Zhang and Kai-Kit Wong "User Association in Massive MIMO and mmWave Enabled HetNets Powered by Renewable Energy" *Wireless Communications and Networking Conference (WCNC), IEEE,* September, 2016.

[14] Dr. Emil Bjrnson "Massive MIMO for 5G" *Tutorial at 2015 IEEE International Conference on Communications ICC 2015* June 8, London, UK.

[15] Kan Zheng, Suling Ou and Xuefeng Yin "Massive MIMO Channel Models: A Survey" *International Journal of Antennas and Propagation* Volume 2014 (2014), Article ID 848071, June, 2014.

[16] Saeed Kaviani and Witold A. Krzymien "On the Optimality of Multiuser Zero-Forcing Precoding in MIMO Broadcast Channels" *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th* June, 2009.