Buffer-Aided 5G Cooperative Networks: Considering the Source Delay

Mohammad Alkhawatrah Wolfson School Loughborough University Loughborough, UK m.alkhawatrah@lboro.ac.uk Yu Gong Wolfson School Loughborough University Loughborough, UK y.gong@lboro.ac.uk Omar Aldabbas Faculty of Engineering AlBalqa' Applied University Amman, Jordan o.aldabbas@bau.edu.jo

Abstract-Applying relays that employed with data buffers drastically enhance the performance of cooperative networks. However, this enhancement faces some difficulties, the biggest obstacle is lengthening the packet delay. In this paper, a new factor the source delay affecting the packet delay, which was not considered in previous studies, is thoroughly studied. The importance of this, is making the calculation of the total delay that messages encounter before reaching their destination more accurate, this is crucial especially in applications that require their messages to get transmitted as fast as possible. Markov chain is employed to model the system and analyze the source delay. Numerical simulations verify the analytical model, and it shows the massive impact of the source delay. The results show that buffer-aided relays can beat non-buffer relays in terms of average packet delay in some cases especially at low SNR range, this makes adding buffers to relays more attractive solution in 5G applications.

Index Terms—buffer-aided relays, cooperative networks, source delay, 5G

I. INTRODUCTION

Relay selection increases the degree of freedom that the network has in selecting, thus it improves many performance metrics like diversity gain of cooperative networks [1]. In particular, many selection schemes are suggested in the literature. One of the traditional schemes is max-min, since each relay has two links (source-relay and relay-destination links) this scheme selects the relay whose weak link (the link which has the lower channel gain between the two links) is the best among the weak links of all relays, then the source transmits the data packet to the selected relay in one time slot. After that, the relay transmits the data packet to the destination in he next time slot, which means no relay can be selected unless both of its source-relay and relay-destination links can handle the transmission, this selecting scheme achieves a full diversity order. After introducing data buffers to relays, the constraint in maxmin, that both of source-relay and relay-destination links have to be able to transmit is relaxed. After this enhancement by introducing buffers, several selecting scheme is suggested, one of the most attractive scheme is max-link, in this scheme the available link which has

the highest channel gain is selected, as a result this scheme can achieve twice the diversity order that maxmin gives, which is twice the full diversity order [2]-[4].

Nonetheless, the enhancement that adding buffers provides in some of the system performance metrics like the diversity order costs the system to experience higher average packet delay. Specifically, in the non-buffer case max-min, each data packet in the network needs two time slots to be transmitted from the source to the selected relay and then from the relay to the destination. On the other hand, in buffer-aided arrangement maxlink, the compulsion that each data packet has to get to the destination with in two time slots is released, this happens because as the data packet transmitted from the source to the relay, it is stored in the relay's buffer until the relay-destination link of that relay becomes the best link. This mean that each data packet may encounter different values of delay [5].

In buffer-aided schemes (like max-link), the average packet delay linearly proportional to the number of relays and the size of the buffers [6]. In other words, the average packet delay increased as the number of relays or the buffer sizes get increased. While the increment in the number of relays or the buffer sizes worsen the system delay, it enhances some other performance metrics such as outage probability and throughput. Therefore, a trade off between the two sides the delay and the other mentioned metrics has to be considered, however, the new 5G system has extremely low tolerance to high values of the delay, so it operates with very low values of delay, this causes that the buffer-aided relays to be inadequate for the new applications [7], [8].

Reducing the packet delay became a trend in the latest studies related to buffer-aided relay cooperative networks, one of the studies suggested to use adaptive link selection with infinite buffer size [9], however, infinite buffer size is impractical. In [7], the authors has suggested a novel term the target length which controls the link selection, in particular, setting the target length to be as short as possible i.e. target = 0 gives higher priority to relay-destination links which guarantees that

all packet will leave their buffers as soon as possible. This reduces the packet delay of buffer-aided schemes, yet the delay that each packet faces at the source is not considered, based on that the traditional max-min greatly outperformed the buffer-aided systems in term of packet delay. In this paper, we propose considering the delay which each packet encounter at the source node, this makes the comparison among the above mentioned schemes unbiased and closer to be a fair comparison.

The main contributions in this paper are listed as follows.

- 1) Considering the source delay in buffer-aided relays network which gives more accurate results.
- 2) Deriving the analytical expressions for the bufferaided network delay based on the new definition of the delay where the source delay is considered. The analysis done based on modeling the system with Markov chain.

The remainder of this paper is organised as follows. In Section II the system model is presented. Section III discuses the outage probability in buffer-aided systems. While section IV present the average packet delay analysis for buffer-aided networks. In Section V, the asymptotic performance of the system is discussed. Simulation results demonstrating the comparison buffer-aide and non-buffer system is VI. Finally, summary concludes this paper in Section VII.

II. SYSTEM MODEL

The system model of the buffer-aided relay networks is shown in Fig. 1, where there are one source node S, K half-duplex decode-and-forward (DF) relay nodes denoted as R_k , $k = 1, \dots, K$ and a destination D. The channel coefficients for $S \to R_k$, $R_k \to D$ links are denoted as h_{s-r_k} , h_{r_k-d} respectively. All channels have flat Rayleigh fading coefficient that remain constant within the time slot and change independently from one slot to another. Every relay R_k is equipped with one L-size buffers for data transmissions to D. We assume that source always has enough information to send to relays in all time slots, in other words, an infinite backlog at source node is assumed. In addition, we assume that the source and the destination are not directly connected. Without losing generality, we assume unity transmit powers P_t at all transmit nodes, same assumption for the noise variances at all receiving nodes σ^2 . The average channel gains for $S - R_k$ link is $E[h_{s-r_k}(t)]$ and for $R_k - D$ link is $E[h_{r_k-d}(t)]$, where E[.] is the expectation. The average channel SNRs are

$$\bar{\gamma}_{s-r_k} = E[|h_{s-r_k}(t)|^2], \ \bar{\gamma}_{r_k-d} = E[|h_{r_k-d}(t)|^2],$$
 (1)

where E[.] is the expectation.

Before considering the source delay, we discuss the available schemes. Recently, a new principle of buffer-



Figure 1: System model for the buffer-aided relay network.

aided max-link relay selection has been suggested in [7], the main idea is to give higher priority to relaydestination link over the source-relay link in all relays, this assures the transmission of the packets resides in buffers as soon as possible. This was not the case in earlier versions of max-link, where all links have the same priority and the link with the highest channel gain is selected.

As the data packets face a longer queues in buffers the delay that each packet encounters becomes larger. Therefore, the average packet delay increases linearly with number of relays and buffers sizes. This is the case in the max-link scheme, for example, let k be the number of relays with buffer size $L \ge 1$. All buffers are assumed to be initially empty, then the packet x_1 is sent to the relay R_1 at time t = 1. After that, at the next time t = 2, all buffers are empty except for R_1 because it contains x_1 . Hence, totally there are (K + 1) available links to be selected: K source-relay links, because all $(S - R_k)$ links are available even when the buffers are empty, and one relay-destination $(R_1 - D)$ link is available as R_1 buffer is the only buffer which is not empty.

Since the max-link scheme consistently chooses the best link among all the available links, the probability that $R_1 - D$ is chosen and x_1 is transmitted to the destination is 1/(K + 1). Which means that it is more likely that x_1 stays in R_1 buffer at t = 2 with the probability of K/(K+1), this adds one extra time slot to the packet delay. It is obvious that this extra delay may be avoided by sending x_1 to the destination directly at t = 2, if the corresponding $R_1 - D$ link is not in outage, even if it is not the best link. This practically explains the mean of giving higher priority to chose the $R_k - D$ links; while $S - R_k$ only chosen when no $R_k - D$ link can be chosen. As a result, the lengths of packet queuing at all buffers are minimized, similarly the average packet delay is minimized.

In summary, at time slot t, the link is chosen as follows:

1) Choose the link which has the highest channel

gain $(|h_{\tau_k d}|^2)$ among all the available $R_k - D$ links. If the nominated link is not in outage, the corresponding relay transmits a packet to the destination from its buffer.

2) If the selected link is in outage at time t then all other $R_k - D$ links are also in outage. In this case, the scheme selects the link with the highest channel gain $(|h_{sr_k}|^2)$ among all the available (the corresponding buffer is not full) $S - R_k$ links $(|h_{sr_k}|^2)$. If the chosen link is up, the source sends a packet to the corresponding relay then the packet is reserved in its buffer. Otherwise, the system is in outage.

Based on [7], the above discussed scheme achieved the best results among the available buffer-aided relay selecting schemes proposed in the literature in terms of average packet delay. This is the reason why the mentioned selection scheme is employed in this study. However, the previous studies have not considered the delay at the source S (source delay), which buffers help in reducing its value. In particular, source delay is proportional to the outage probability in the system, and buffer-aided systems outperform non-buffer systems in terms of outage probability. This may be considered as a privilege to employing buffers with relays, yet no study, to the best of our knowledge, has considered the source delay. As a result, considering source delay shows that max-link can beat max-min in terms of delay in some cases based on the discussion through this chapter.

In fact considering source delay is easy to be implemented since it does not require any additional information than that in the existing buffer-aided max–min scheme. In the following two sections, outage and delay performance of the proposed scheme will be analyzed, respectively.

III. OUTAGE PROBABILITY

Based on Shannon's law, the link capacity is

$$C_{p_k}(t) = \log_2 (1 + \gamma_{p_k}(t)),$$

$$p_k \in \{s - r_k, r_k - d\}, \ k = 1, \cdots, K,$$
(2)

outage occurs if the link capacity is less than the target data rate

$$P\{\log_2\left(1+\gamma_{p_k}(t)\right) < \eta\} = 1 - e^{\left(-\frac{2\eta}{\gamma_{p_k}}\right)}$$
(3)

where C and η are the capacity and the target data rate respectively.

In each of the relays buffers, the numbers of data packets represent a state. Since there are K relays with L buffer size, there are $(L + 1)^K$ states in total. Every state suggests the numbers of the available $S - R_k$ and $R_k - D$ links. Any $S - R_k$ link is considered available if the buffer belongs to the corresponding relay has the ability to receive a new packet (not full), and any $R_k - D$ link is available if the corresponding relay buffer has

packets to be transmitted (not empty). The *l*-th state vector is defined as

$$\mathbf{q}^{(l)} = [q_1^{(l)}, q_2^{(l)}, \cdots, q_K^{(l)}], \qquad l = 1, \cdots, (L+1)^K,$$
(4)

where $q_k^{(l)}$ is the buffer length at R_k at state $\mathbf{q}^{(l)}$

By taking all possible states into consideration, the outage probability of the proposed max-link scheme can be obtained as

$$P_{out} = \sum_{i=1}^{(L+1)^K} P_{out}^{\mathbf{q}(i)} \pi_i.$$
 (5)

where π_i is the stationary probability for state $\mathbf{q}(i)$, and $P_{out}^{\mathbf{q}(i)}$ is the outage probability at state $\mathbf{q}(i)$. Particularly, for independent and identically distributed (i.i.d) Rayleigh fading channels, the instantaneous SNR of both channels, $[h_{s-r_k}, h_{r_k-d}]$, in each relay is independently exponentially distributed. Furthermore, in the employed max-link scheme, outage occurs if all $R_k - D$ links then all $S - R_k$ links are in outage. Hence, the outage probability at state \mathbf{q}_1 is given by

$$P_{out} = p_{sr1}.p_{r1d} \tag{6}$$

where

$$p_{sr1} = (1 - \exp^{-\frac{2^{\eta} - 1}{\bar{\gamma}_{s-r_1}}})$$
(7)

$$p_{r1d} = (1 - \exp^{-\frac{2^{\eta} - 1}{\bar{\gamma}_{r_1 - d}}})$$
 (8)

where p_{sr} and p_{rd} are the probabilities that all available $S-R_k$ links and R_k-D links are in outage, respectively.

Regarding stationary probability, buffer states can be modeled as a discrete time Markov chain, the transition matrix of the Markov chain is denoted as **A** representing $(L+1)^K * (L+1)^K$ state transition, **A**_{mn} is the notation for the *m*th row and *n*th column entry, which expresses the transition probability to move from state q_n at time *t* to state q_m at time t + 1:

$$\mathbf{A}_{mn} = P(X_{t+1} = q_m | X_t = q_n)$$
(9)

The described Markov chain with the transition matrix **A** has two properties: irreducible and aperiodic. The Markov chain is considered irreducible if all states are reachable by all other states in the chain, and if the probability of staying at any state higher than zero, then the Markov chain is aperiodic, see [10], [11]. In irreducible and aperiodic Markov chain, the stationary state probability vector is obtained as

$$\boldsymbol{\pi} = (\mathbf{A} - \mathbf{I} + \mathbf{B})^{-1} \mathbf{b}, \tag{10}$$

where $\boldsymbol{\pi} = [\pi_1, \pi_2, \cdots, \pi_{(L+1)}], \pi_m$ is the probability that the buffer state is $\mathbf{q}_m, \mathbf{b} = [1, \cdots, 1]^T$, **I** is the notation of the identity matrix and **B** denotes an (L + $1)\times (L+1)$ matrix with all elements have the value of one.

The definition of the outage probability of the system is the probability that all relays neither transmit to the destination nor receive from the source. When this happens, the number of packets reside in the buffers remains the same, this means the Markov chain remains in the same state, so outage probability can be obtained as follows:

$$P_{out} = \sum_{i=1}^{(L+1)^K} \pi_i \mathbf{A}_{ii}$$
(11)

where A_{ii} is the diagonal elements of A.

IV. AVERAGE PACKET DELAY

The traditional definition of the delay of a packet in the buffer-aided schem is the duration between the time the packet leaves the source node and the time it arrives the destination. However, in this paper we are adding to the described delay definition the time that the packet has to wait at the source when it is ready to be transmitted. The average packet delay (including the delay at the source) of the system is given by

$$\bar{D} = \bar{D_s} + \bar{D_{sr}} + \bar{D_r} \tag{12}$$

where \bar{D}_s denotes the source delay, \bar{D}_{sr} is the delay caused by transmitting data packet from the source to a relay and \bar{D}_r is the delay at each relay node.

Firstly, since \overline{D}_s proportional to the outage probability P_{out} , we need to calculate P_{out} . As presented in III, P_{out} is obtained through Markov chain model, let L = 3 then the Markov chain which models the system is shown in Fig. 2.



Figure 2: Markov chain for L = 3 system.

The transition matrix of the Markov chain is

$$\mathbf{A} = \begin{bmatrix} p_{sr} & \tilde{p_{sr}} & 0 & 0\\ \tilde{p_{rd}} & p_{sr}p_{rd} & \tilde{p_{sr}}p_{rd} & 0\\ 0 & \tilde{p_{rd}} & p_{sr}p_{rd} & \tilde{p_{sr}}p_{rd}\\ 0 & 0 & \tilde{p_{rd}} & p_{rd} \end{bmatrix}$$

based on that, the stationary state probabilities can be calculated by (10), hence

$$P_{out} = \sum_{i=0}^{3} \pi_i \mathbf{A}_{ii}.$$
 (13)

Since the assumption is that the source has infinite data to be transmitted, source delay considered only at empty buffer, because when any buffer has a data packet then the delay is calculated in the delay packets encounter at the relay, thus

$$\bar{D_s} = \bar{P_{out}} \pi_0 \sum_{i=0}^n P_{out}^i \tag{14}$$

as $n \to \infty$:

$$\bar{D}_s = \bar{P_{out}} \pi_0 \left(\frac{1}{1 - P_{out}}\right) = \pi_0 \tag{15}$$

where P_{out}^{-} is the probability of no outage and n is source delay instant.

Secondly, as it takes one time slot to send a packet from the source to a relay node, so the average D_{sr} is one. Finally, since the average delay in every relay is the same, the average delay in one relay R_k only is analyzed. Based on Little's law, the average packet delay at relay R_k is given by

$$\bar{D_{rk}} = \frac{\bar{L_{rk}}}{\bar{\eta_{rk}}} \tag{16}$$

where L_{rk}^{-} and η_{rk}^{-} are the notations of average queuing length and average throughput at the relay R_k respectively. The average of the queuing length at R_k is calculated by averaging the queuing lengths at the relay R_k 's buffer (denoted as Q_k) over all states

$$\bar{L_{rk}} = \sum_{i=1}^{(L+1)^K} \pi_i q_k^{(i)}$$
(17)

On the other hand, because selecting any of the relays has the same probability, the average throughput at the relay R_k is given by

$$\bar{\eta_{rk}} = \frac{\bar{\eta}}{K} \tag{18}$$

where $\bar{\eta}$ is the average throughput of the overall network. For delay-limited transmission (the capacity is calculated when the outages is not tolerated) as in [12] and [13], the average throughput $\bar{\eta}$ is obtained as

$$\bar{\eta} = R(1 - P_{out}) \tag{19}$$

where R is the average data rate of the system . In the max-link scheme, every packet needs two time slots (do not have to be consecutive) to reach the destination, we have R = 1/2, and thus

$$\bar{\eta_{rk}} = \frac{(1 - P_{out})}{2K} \tag{20}$$

substituting (17) and (20) into (16) gives

$$\bar{D}_r = \frac{2K(\sum_{i=1}^{(L+1)^K} \pi_i)q_k^{(i)}}{1 - P_{out}}.$$
(21)

V. Asymptotic performance

This section studies the asymptotic performance of the max-link scheme in the presence of source delay when the average channels SNR for both of source-relay and relay-destination links goes to infinity. Accordingly

$$\lim_{E[h_{sr}]\to\infty} p_{sr} = 0 \tag{22}$$

it is clear that outage (caused by the link state) is impossible to occur when SNR is high enough, since the source delay only occurs when $S - R_k$ links being in outage, the source delay approaches zero as SNR goes to infinity

$$\lim_{E[h_{sr}]\to\infty}\bar{D_s}=0.$$
(23)

Assume that all the buffers at time t are empty, so that the system is in the state $q_k^{(0)}$. At this point, a packet will be sent to a relay at time (t+1), and the system moves to state $q_k^{(1)}$. After that, the packet in the buffer need to be transmitted to the destination at (t+2), then the system returns to state $q_k^{(0)}$. And this process continues, thus

$$P(q_k^{(0)}) = P(q_k^{(1)}) = \frac{1}{2K}$$
(24)

$$\lim_{E[h_{sr}]\to\infty} \bar{L_{rk}} = 0.P(q_k^{(0)}) + 1.P(q_k^{(1)}) = \frac{1}{2K}$$
(25)

$$\lim_{E[h_{sr}] \to \infty} \eta_{rk}^{-} = \frac{\lim_{E[h_{sr}] \to \infty} (1 - P_{out})}{2K} = \frac{1}{2K} \quad (26)$$

from (16), we find $\lim_{E[h_{sr}]\to\infty} \bar{D_r} = 1$, in result

$$\lim_{E[h_{sr}] \to \infty} \bar{D} = 0 + 1 + 1 = 2$$
 (27)

this means, at high values of SNR, the delays in bufferaided and non-buffer (max-min) cases are the same. This will be verified in the next section.

VI. NUMERICAL SIMULATIONS

In all simulations below, the target transmission rate is set to $\eta = 2$ bps/Hz, the buffer size is set to L = 5 for every buffer and all noise powers σ^2 are normalized to unity. The average channel gains are set to $E[h_{sr}] = E[h_{r_kd}] = 0.5$.

In Fig. 3, we show the average packet delay vs SNR (dB) in single relay system. It is worth noting that the link selection based on prioritizing the relay-destination link, which meas that the target is to keep the buffer empty as long as possible i.e. target = 0. The simulation results clearly verifies the results derived in previous



Figure 3: Source delay (\overline{D}_s) only in single relay system theoretical *vs* simulation.

sections. In other words, \overline{D}_s starts high and then goes to zero as the SNR reaches high values.

Fig. 4 and Fig. 5 show the impact of considering the source delay (\bar{D}_s) on the average packet delay. It can be clearly seen that prior considering \bar{D}_s , delay in max-min is not beatable, however, after considering \bar{D}_s , which makes the comparison fairer, buffer-aided system outperforms non-buffer system in some cases like in target = 0.

As stated earlier, adding more relays boosts the system delay in buffer-aided systems, this is shown in Fig. 6. This leads to non-buffer system to beat the buffer-aided systems in term of average packet delay. However, the difference is still insignificant oppose to the case of \bar{D}_s being unconsidered. It is worth noting that adding more relays reduces non-buffer total system delay.

Interestingly, when we applied the broadcast technique, which means transmitting each packet to all relays all at once simultaneously, the delay in both cases get closer. Furthermore, in Fig. 7, when we reduced the buffer size to L = 1 while employing broadcasting, the buffer-aided three relay archives lower delay than in non-buffer case. This techniques and others may be studied in details in future research.

VII. CONCLUSION

Longer delay is than main struggle in applying relays. Therefore, the impact of considering source delay that each packet encounters before being transmitted is studied in this paper, this makes the comparison more accurate. The presented results show that the total delay of target = 0 max-link is lower than the total delay of max-min in single relay network. In other words, buffer case might have lower delay than the non-buffer case. In addition, max-link beats max-min in three relays network with considering some techniques like broadcasting.



Figure 4: Average packet delay in single relay system without considering the source delay (\overline{D}_s) .



Figure 5: Average packet delay in single relay system with \bar{D}_s .

REFERENCES

- N. Nomikos et al., "A Survey on Buffer-Aided Relay Selection," in IEEE Communications Surveys and Tutorials, vol. 18, no. 2, pp. 1073-1097, Secondquarter 2016.
- [2] A. Ikhlef, D. S. Michalopoulos and R. Schober, "Buffers Improve the Performance of Relay Selection," 2011 IEEE Global Telecommunications Conference - GLOBECOM 2011, Kathmandu, 2011, pp. 1-6.
- [3] N. Zlatanov, R. Schober and P. Popovski, "Throughput and Diversity Gain of Buffer-Aided Relaying," 2011 IEEE Global Telecommunications Conference - GLOBECOM 2011, Kathmandu, 2011, pp. 1-6.
- [4] N. Nomikos, T. Charalambous, D. Vouyioukas, G. K. Karagiannidis and R. Wichman, "Relay Selection for Buffer-Aided Non-Orthogonal Multiple Access Networks," 2017 IEEE Globecom Workshops (GC Wkshps), Singapore, 2017, pp. 1-6.
- [5] D. Q. Qiao and M. C. Gursoy, "Buffer-Aided Relay Systems under Delay Constraints: Potentials and Challenges," in IEEE Communications Magazine, vol. 55, no. 9, pp. 168-174, Sept. 2017.
- [6] P. Xu, Z. Ding, I. Krikidis and X. Dai, "Achieving Optimal Diversity Gain in Buffer-Aided Relay Networks With Small Buffer Size," in IEEE Transactions on Vehicular Technology, vol. 65, no. 10, pp. 8788-8794, Oct. 2016.



Figure 6: Average packet delay in three relays system with \bar{D}_s .



Figure 7: Broadcast impact on average packet delay in three relays system with \overline{D}_s and L = 1.

- [7] Z. Tian, Y. Gong, G. Chen and J. A. Chambers, "Buffer-Aided Relay Selection With Reduced Packet Delay in Cooperative Networks," in IEEE Transactions on Vehicular Technology, vol. 66, no. 3, pp. 2567-2575, March 2017.
- [8] V. Jamali, N. Zlatanov and R. Schober, "Bidirectional Buffer-Aided Relay Networks With Fixed Rate Transmission—Part II: Delay-Constrained Case," in IEEE Transactions on Wireless Communications, vol. 14, no. 3, pp. 1339-1355, March 2015.
- [9] Z. Ding, Y. Gong, T. Ratnarajah and C. F. N. Cowan, "Opportunistic Cooperative Diversity Protocols for Wireless Networks," 2007 IEEE Information Theory Workshop on Information Theory for Wireless Networks, Solstrand, 2007, pp. 1-5.
- [10] Norris, James R, "Markov chains," Cambridge University Press., 1998.
- [11] Berman, Abraham, and Robert J. Plemmons. Nonnegative matrices in the mathematical sciences. Vol. 9. Siam, 1994.
- [12] N. Ahmed, M. A. Khojestapour and R. G. Baraniuk, "Delaylimited throughput maximization for fading channels using rate and power control," IEEE Global Telecommunications Conference, 2004. GLOBECOM '04., Dallas, TX, 2004, pp. 3459-3463 Vol.6.
- [13] R. Narasimhan, "Throughput-Delay Performance of Half-Duplex Hybrid-ARQ Relay Channels," 2008 IEEE International Conference on Communications, Beijing, 2008, pp. 986-990.