



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

An Analysis on Wireless Edge Caching in In-band Full-Duplex FR2-IAB Networks

Citation for published version:

Zhang, T, Biswas, S & Ratnarajah, T 2020, 'An Analysis on Wireless Edge Caching in In-band Full-Duplex FR2-IAB Networks', *IEEE Access*, vol. 8, pp. 164987 - 165002.
<https://doi.org/10.1109/ACCESS.2020.3022725>

Digital Object Identifier (DOI):

[10.1109/ACCESS.2020.3022725](https://doi.org/10.1109/ACCESS.2020.3022725)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

IEEE Access

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Received August 15, 2020, accepted August 28, 2020, date of publication September 8, 2020, date of current version September 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3022725

An Analysis on Wireless Edge Caching in In-Band Full-Duplex FR2-IAB Networks

TONG ZHANG¹, SUDIP BISWAS², (Member, IEEE), AND
THARMALINGAM RATNARAJAH¹, (Senior Member, IEEE)

¹School of Engineering, Institute for Digital Communications, The University of Edinburgh, Edinburgh EH8 9YL, U.K.

²Department of Electronics and Communications Engineering, Indian Institute of Information Technology Guwahati, Guwahati 781015, India

Corresponding author: Sudip Biswas (sudip.biswas@ieee.org)

The work of Tong Zhang and Tharmalingam Ratnarajah was supported in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P009549/1, in part by the U.K.-India Education and Research Initiative Thematic Partnerships under Grant DST UKIERI-2016-17-0060, and in part by the SPARC Project 148. The work of Sudip Biswas was supported by Indian Institute of Information Technology (IIIT) Guwahati through TEQUIP-III of Govt. of India.

ABSTRACT This article develops a 3GPP inspired analytical framework for cache-enabled in-band full-duplex (IBFD) integrated access and backhaul (IAB) heterogeneous network in the FR2 (millimeter wave) band. In particular, by assuming wide-band FR2 channel model, we analyze the performance of cache-enabled IAB nodes that serve multiple users in access in IBFD mode through hybrid beamforming architectures. Since, IBFD transmission is considered, the IAB nodes are able to retrieve any non-cached files simultaneously through backhaul from the IAB donor nodes, which also use hybrid beamforming architectures. Utilizing stochastic geometry tools, we first derive the lower-bound for the average success probability (ASP) of file delivery for the considered system model, using which we evaluate two quintessential performance metrics: latency and average throughput of file delivery under a cache-enabled content delivery phase. Further, as a baseline for comparison, we also consider conventional half-duplex techniques, namely time division duplex and frequency division duplex transmission and evaluate their performance with respect to cache size, node density, content popularity, antenna numbers, blockage density, and residual self-interference. Numerical results demonstrate the precedence of using wireless edge caching as a cost-effective measure for improving the FR2-IAB network's performance. More importantly, we provide several design insights for the implementation of wireless edge caching in IBFD FR2-IAB networks, with respect to required residual self interference cancellation, increasing antenna number, increasing target data rate and IAB node density.


INDEX TERMS Wireless edge caching, in-band full-duplex, FR2, 5G NR, stochastic geometry, integrated access and backhaul networks.

I. INTRODUCTION

In order to meet the exponential rise in demand for wireless data, the orientation of the fifth generation (5G) & beyond technologies can be mainly identified as: i) network densification through small cells [1], ii) moving to frequency range 2 (FR2 *i.e.*, millimeter wave (mmWave)) spectrum (>24.25 GHz) bands [2], and iii) improving spectral efficiency through simultaneous transmission and reception at the same time and frequency bands through in-band full-duplex (IBFD) radios [3]. In particular, although

FR2 provides large chunks of unused spectrum for high data rates, it is quite vulnerable due to weak diffraction ability and susceptible to blockages due to lower wavelength. However, recent channel measurements have demonstrated the feasibility of mmWave transmission for the access link [4]. Further, system level evaluation has proved that considerable benefits on the coverage probability and spectral efficiency can be achieved by densely deploying mmWave-aided small cells [5], which is one of the key driving technology for the evolution of 5G new radio (NR) networks.

Nevertheless, the entire data traffic from the access links is inevitably forwarded to the core network through backhaul networks, which makes the network performance to be highly

The associate editor coordinating the review of this manuscript and approving it for publication was Xingwang Li .

related to the backhaul links. In fact, the increasing demands for higher quality of user experience requires data rate in the Gigabits per second (Gbps) range on the backhaul. Although optical fibre is less likely to be a concern with regards to the backhaul solution for macro cells access to the core network, it is an enormous challenge to bring fibre connection to each and every small cells in a densely deployed heterogeneous network due to both scalability issues and installation costs [6]. Motivated by this, the 3GPP has proposed a novel integrated access and backhaul (IAB) framework within the FR2 frequencies to achieve high spectral efficiency while also supporting reliable transmission [7]. On one hand, with the vast availability of additional bandwidth in FR2, it is practical to achieve data rates of several Gbps for backhaul links while guaranteeing sufficient bandwidth for access links. On the other hand, such a novel FR2-IAB infrastructure recognized by 3GPP is a cost-effective alternative to the wired backhaul that empowers both access and backhaul communications to utilize the same 5G NR technology and share the same resource pool to simplify the overall design complexity in pursuit of higher wireless capacity. Despite the consensus about the ability of IAB to reduce costs, designing an efficient and high-performance IAB network is still an open research challenge. In particular, the dynamic resource allocation between the backhaul and access links is the main technical problems for the FR2-IAB framework. With the wide availability of the FR2 spectrum, it is necessary to investigate how feasible it is to include wireless backhaul communications in the same spectrum as wireless access since recent studies have showed that it plays a significant role in the performance of rate coverage [8], cell edge throughput [9], and sum rate [10].

Additionally, a substantial amount of redundant transmission over networks is generated by users by asynchronously and repeatedly requesting some popular files [11]. In order to deal with the high wastage of resources and aggravation of traffic burden on backhaul links, *wireless edge caching* (WEC), defined as pre-fetching some popular contents in the local caches of base stations (BSs) closer to users, has been proposed as a potential solution. In fact, WEC has the ability to alleviate network backhaul load, reduce content delivery latency due to shortened communication distance, and improve the overall throughput [12]. Accordingly, caching techniques for wireless networks have gained significant attention with many undergoing studies of late on the design and analysis of cache-enabled wireless networks. In this regard, femto caching was considered in [13], where the optimal caching placement through distributed caching helpers with low-rate backhaul capacity, but high storage capacity was proposed to minimize the expected downloading delay. However, the considered network model did not capture the stochastic nature of channel fading and geometric locations of BSs and users. Similarly, a stochastic geometry aided cache-enabled small cell network was analyzed with respect to the outage probability and average file delivery rate in [14]. Further, leveraging the analysis

in [14], a cache-enabled heterogeneous networks (HetNets) consisting of a tier of multi-antenna macro BSs (MBSs) overlaid with a tier of caching helpers was considered in [15], where the optimal caching placement was analyzed in terms of probability of success of file delivery and the area spectral efficiency (ASE). However, none of these works were validated for the FR2 spectrum. Nevertheless, WEC in mmWave D2D communications and massive multi-input multi-output (MIMO)-aided HetNets were investigated in [16] and [17], respectively. In [18], WEC was considered for a hybrid FR1 (<7.225 GHz) and mmWave HetNet, where optimal caching placement schemes with respect to the average success probability (ASP) of file delivery was evaluated.

Furthermore, analogous to conventional wireless networks, existing architectures for IAB networks employ half-duplex (HD) radios only, which cannot simultaneously utilize the time and frequency resources of a spectrum, thereby not only reducing the spectrum efficiency, but also increasing latency in communication. This problem can be mitigated by implementing IBFD radios. Recent advances in various self-interference (SI) cancellation techniques in antenna, analog and digital domains [19], [20] invalidate a long-held assumption that transmission and reception cannot happen simultaneously within the same time-frequency band. In particular, recent works such as [21]–[24] justify the use of IBFD radios in future wireless communications. With separate antennas for transmission/reception, IBFD-MIMO radios, can support bidirectional communication, thus improving the performance of the IAB network further. However, unlike the above cited literature, the primary goal of this article is to analyse the feasibility and performance of a cache enabled IBFD IAB network considering a wide-band FR2 channel model.

Based on the above, we would like to note that very few works to date have taken into consideration the IBFD operation for cache-enabled FR2-IAB networks. In fact, 3GPP clearly puts forward the requirement for the spectral efficiency enhancements [7] for IAB through the application of innovative technologies, which is why we consider the innovative framework of IBFD FR2-IAB network in conjunction with WEC. In particular, we explore a more ambitious solution for FR2-IAB networks than what is currently available in literature by i) utilizing the same spectrum resources and infrastructure of the FR2 transmission band to serve cellular users in access as well as the IAB nodes in backhaul, and ii) incorporating WEC by equipping the IAB nodes with inexpensive memory. The primary contributions of this work are summarized as follows:

- Unlike existing out of band IAB HD solutions, in this work we consider a WEC-enabled FR2-IAB architecture operating in an IBFD mode, such that in the event of a cache-miss scenario at the IAB nodes it is possible to retrieve the non-cached files from IAB donor nodes with the least possible latency and minimum degradation in data rate, thus theoretically providing twice the throughput of baseline HD solutions.

- Unlike existing literature on IAB, this work does not require bandwidth partitioning due to the IAB nodes operating in IBFD mode, whereby the whole bandwidth allocated to an IAB node can be utilized for both access and backhaul at the same time.
- Unlike other FR2 narrow band channel models, this work considers wide-band FR2 channels that introduces frequency selectivity, whereby the impact of multi-path delay spread is more significant. In order to cope with frequency selectivity and simultaneously enabling IBFD operation at IAB nodes in both access and backhaul links, we consider that each IAB node (IAB donor node) simultaneously serves multiple users (IAB nodes) through fully-connected OFDM-based hybrid beamforming architectures.
- By utilizing stochastic geometric tools, that accounts for interference among the operating IAB donor nodes and IAB nodes, uncertainties both in node locations, channel fading (*i.e.*, path loss), and cell loads, we derive tractable signal-to-interference-and-noise ratio (SINR) expressions at the receiver node located at the origin. Next, considering the effects of blockages in FR2 transmission, both line-of-sight (LOS) and non-LOS (NLOS) transmissions are separated from the original point process and then analyzed.
- In order to simplify the expressions and obtain tractable results, we first define the path loss process, with which we derive the user association probability, the statistics of the serving distance, and mean cell loads. Next, using these expressions we derive the lower bounded ASP of file delivery, average throughput, and the latency of file delivery for the IBFD case with WEC in access. In addition, for the sake of comparison we also consider baseline HD modes (time division duplex (TDD) and frequency division duplex (FDD)) at the IAB nodes and the no caching scenario where the IAB nodes are not equipped with any storage memory.
- Numerical results show that WEC is an effective solution to improve the network performance of future IAB networks albeit certain tradeoffs. Compared to TDD mode, IBFD is superior with respect to ASP, latency, and throughput of file delivery. However, while FDD mode performs well with respect to rate coverage, IBFD outperforms it with respect to latency and throughput. In particular, we observe a tradeoff in performance between FDD and IBFD modes with respect to cache size. Apart from cache size, we also evaluate the network performance with respect to node density, content popularity, antenna numbers, blockage density, and residual SI cancellation (RSIC) factor. One of the significant results that we observed is that WEC has the ability to reduce SI cancellation requirements for the deployment of practical IBFD FR2-IAB networks.

II. SYSTEM MODEL

In this section, we describe the cache-enabled IBFD FR2-IAB network. Some of the important notations used in the paper are summarized in Table 1.

A. NETWORK TOPOLOGY

We consider the downlink of an IAB network as depicted in Fig. 1, where IAB donor nodes co-exist with IAB nodes and are distributed in the 2-D Euclidean plane \mathbb{R}^2 according to two independent homogeneous Poisson point processes (PPPs) Φ_m and Φ_s with densities λ_m and λ_s , respectively [5], [17], [25], [26]. Due to the IBFD operation, same spectral resources under FR2 band is used by the IAB donor nodes to serve IAB nodes in backhaul and IAB nodes to provide access to cellular users, that are assumed to be distributed according to another PPP Φ_u with density λ_u . We assume that there is no direct IAB donor node-user access link, which ensures that the users are only served by the IAB nodes. Nevertheless, this assumption does not restrict the IAB donors from interfering with the users' access links. The propagation among each IAB node and its associated users takes place via a fully connected hybrid precoder that combines both radio frequency (RF) and baseband (BB) precoding. Similarly, each IAB donor node simultaneously serves IAB nodes through another hybrid beamforming architecture. Further, in view of the sparsity of FR2 channels, this work assumes that all scatters take place in the azimuth plane and are uniformly distributed in $[0, 2\pi]$ [27]. Therefore, each IAB donor node, IAB node, and user is assumed to employ a uniform linear array (ULA) of size n_t^m , n_t^s , and n_r^u , respectively¹ [27], ([18], Remark 1). In particular, each IAB donor node and IAB node has n_{RF}^m RF chain and n_{RF}^s RF chains, respectively, for transmitting. Besides, for simplicity we assume that multiple users (IAB nodes) are served via one stream per user (IAB node) by its serving IAB node (IAB donor node). Thus it is sufficient to assume that each user (IAB node) employs a RF-only combiner with the antenna number n_r^u (n_r^s) to decode the transmitted signal as described in [27]. More details on the propagation model will be elaborated in the subsequent section.

According to the Slivnyak's theorem, the analysis henceforth is performed at the typical user, which without loss of generality is assumed to be located at the origin. Further, the IAB nodes are equipped with IBFD radios and are assumed to be equipped with storage memory, such that the cache-miss files are retrieved from IAB donor nodes via wireless backhaul. Since IBFD allows both transmission and reception at the same time and frequency band W , the IAB operation does not require any bandwidth partitioning and hence is a true in-band IAB solution. However, the backhaul links at the IAB nodes are susceptible to SI from the access links. In this work we assume that SI is already cancelled

¹Due to very short wavelength of FR2 signals, it is reasonable to assume that even small mobile user equipments will be equipped with multiple antennas.

TABLE 1. Summary of notations.

Notations	Physical meaning
Φ_m, Φ_s, Φ_u	PPP distributed locations of IAB donor nodes, IAB nodes, and users, respectively
$\lambda_m, \lambda_s, \lambda_u$	Spatial densities of IAB donor nodes, IAB nodes, and users, respectively
P_m, P_s	Transmit power of each IAB donor node and IAB nodes, respectively
n_t^m, n_t^s	Number of transmit antennas at each IAB donor node, IAB node, respectively
n_r^s, n_r^u	Number of receive antennas at each IAB node and user, respectively
n_{RF}^s, n_{RF}^m	Number of RF chains at each IAB node and IAB donor node, respectively
W	Total bandwidth
$\mathcal{F} = \{f_1, \dots, f_F\}$ (i.e., $ \mathcal{F} = F$)	The finite file set with F total files
S	File size in bits
$\mathcal{Q} = \{q_1, \dots, q_F\}$	Content popularity vector with q_i denoting the probability of request of the i th file ($i \in \{1, \dots, F\}$)
ν	Skewness of the content popularity distribution
C	Cache size of each IAB node
$\tilde{\Omega} = \{\tilde{\omega}_1, \dots, \tilde{\omega}_F\}$	Caching probability set with $\tilde{\omega}_i$ denoting the probability that the file i is cached in the IAB node
M_m, M_s	Maximum number of IAB nodes and users served by each IAB donor node and IAB node, respectively
N_x^s, N_x^m	Number of associated users for each IAB node and IAB donor node (located at x), respectively
$\mathcal{U}_x^m, \mathcal{U}_x^s$	Set of scheduled IAB nodes and users served by each IAB donor node and IAB node located at x , respectively
U_x^m, U_x^s	Cardinality of the set \mathcal{U}_x^m and \mathcal{U}_x^s , respectively
β	Blockage density
$\alpha_{\mathcal{L}}, \alpha_{\mathcal{N}}$	LOS and NLOS path loss exponent, respectively
d	Time delay of each path
$\eta_{\mathcal{L}}, \eta_{\mathcal{N}}$	Number of LOS and NLOS scatters, respectively
$\mathbf{a}_R(\cdot), \mathbf{a}_T(\cdot)$	Steering vector responses of the receiver and transmitter, respectively
θ, ϕ	Angles of arrival and departure, respectively
T_s	Sampling time
ϵ	The roll-off factor of the raised-cosine filter
ν	Target rate
D	Cyclic prefix
K	Total number of subcarriers
$\mathcal{A}_s^{\mathcal{L}}, \mathcal{A}_s^{\mathcal{N}}, \mathcal{A}_m^{\mathcal{L}}, \mathcal{A}_m^{\mathcal{N}}$	Association probabilities of each user/IAB node associated with IAB node/IAB donor node in LOS and NLOS transmission, respectively
$\mathcal{P}_s(\nu), \bar{R}, T_{(\cdot)}$	Average success probability, throughput, and latency of file delivery, respectively
σ^2	Noise power

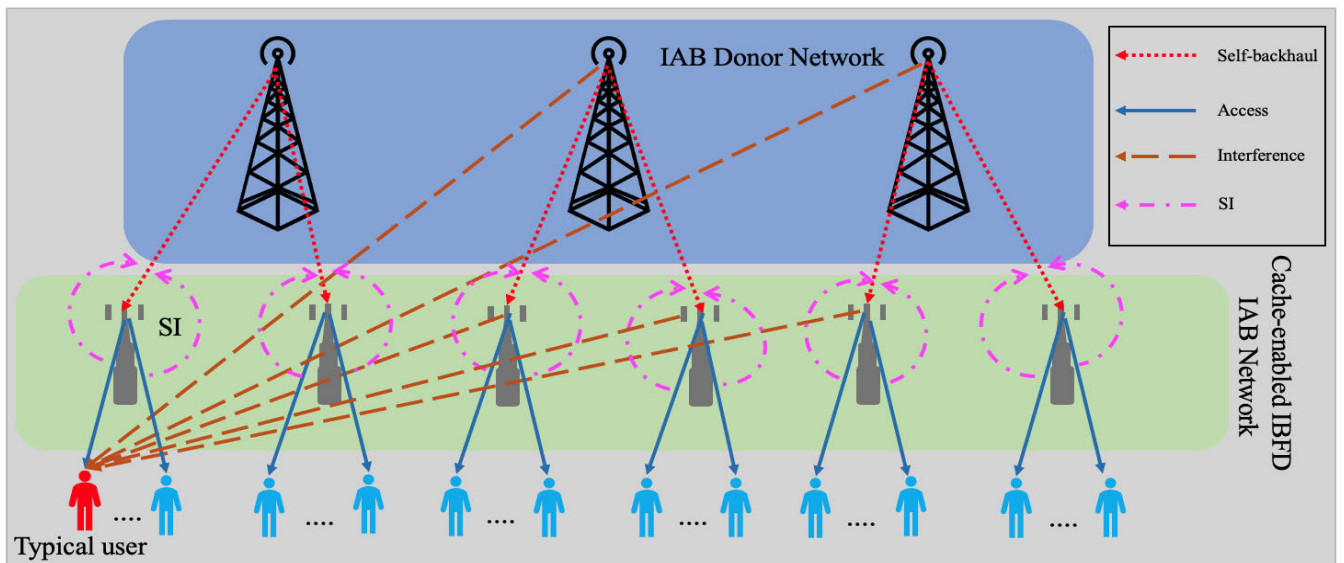


FIGURE 1. An illustration of the considered IBFD FR2-IAB network.

through analog and digital cancellation techniques [19] and only a residual self-interference (RSI) remains in the IAB nodes.

Furthermore, the IAB donor nodes are connected to the core network via high-capacity optical fibre links

and subsequently, it is justifiable to assume that the IAB donor nodes have full access to all files because of high reliability and stability of wired backhaul links. For comparison, we also consider baseline HD IAB solutions, whereby we explore both TDD and FDD modes.

B. WEC MODEL

We consider a finite file set $\mathcal{F} = \{f_1, \dots, f_F\}$, where F is the total number of files and f_i is the i th most popular file with equal file size S bits.² The content popularity distribution vector (*i.e.*, the probability of request of each file in \mathcal{F}) is denoted as $\mathcal{Q} = \{q_1, \dots, q_F\}$, which is assumed to be perfectly known as a prior.³ Each user independently requests the file f_i according to Zipf distribution [18], [28], [29], given as $q_i = (i^\nu \sum_{j=1}^F j^{-\nu})^{-1}$, where ν is the coefficient of skewness that controls the content popularity. Further, the caching capacity of each IAB node is defined as CS bits, where C is the total number of stored files such that $C \leq F$. In particular, this work considers a generic WEC policy in a probabilistic manner, where we define a caching probability set $\tilde{\Omega} = \{\tilde{\omega}_1, \dots, \tilde{\omega}_F\}$, such that $0 \leq \tilde{\omega}_i \leq 1$ due to the probability property and $\sum_{i=1}^F \tilde{\omega}_i \leq C$ due to the caching constraint in an average sense. The motivation behind the use of such a policy is that this makes it analytically tractable to incorporate the stochastic geometric framework to investigate different WEC strategies. In particular, in this work we analyze two commonly used WEC strategies: 1) caching most popular files (MC) and 2) uniform caching (UC). For MC, the caching probability $\tilde{\omega}_i = 1$ for $\forall i \in [1, C]$ and $\tilde{\omega}_i = 0$ for $\forall i > C$. As for UC, the caching probability $\tilde{\omega}_i = \frac{C}{F}$ for $\forall i \in [1, F]$. Based on the caching probabilities, the average cache hit probability (considering that the requested files are stored in the IAB nodes) is given as $p_{\text{hit}} = \sum_{i=1}^F q_i \tilde{\omega}_i$. Finally, for comparison, we also consider the baseline scenario without caching, where the IAB nodes are not equipped with any storage memory and cannot store any files.

C. USER ASSOCIATION AND CELL LOAD

Based on the least path loss user association, every user is associated with an IAB node. Since each IAB node cannot serve more users than its available RF chains in one time-frequency resource block, we assume that the maximum number of users served by an IAB node in each resource block is limited to M_s such that $M_s \leq n_{\text{RF}}^s$, where n_{RF}^s is the number of RF chains of each IAB node. Similarly, each IAB node is associated with an IAB donor node (to retrieve cache miss files) based on least path loss, where each IAB donor node can serve a maximum of M_m associated IAB nodes such that $M_m \leq n_{\text{RF}}^m$. Now, let \mathcal{U}_x^s denote the set of all users in Φ_u which are scheduled by the IAB node at x in one resource block. The cardinality of \mathcal{U}_x^s is expressed as $U_x^s = \min(M_s, N_x^s)$, where N_x^s is the number of associated users of the IAB node at x . Similarly, let \mathcal{U}_x^m denote the set of all IAB nodes in Φ_s which are scheduled by the IAB donor node at x and its cardinality is expressed as $U_x^m = \min(M_m, N_x^m)$, where N_x^m is the number of associated IAB nodes of the IAB donor node at x .

²In the event of unequal file size, each file can be divided into small partitions of the same size, with each partition being treated as an individual file.

³In practice, the content popularity is unknown and should be estimated through some estimation techniques, which is a promising future research topic.

According to [5], which provides the approximated mean cell load for such non-Poisson Voronoi association cells, we have $N_x^s = 1 + 1.28 \frac{\lambda_u}{\lambda_s}$ for the tagged IAB node and $N_x^s = \frac{\lambda_u}{\lambda_s}$ for the non-tagged IAB node. This mean number of associated users per cell includes both cache-hit and cache-miss users. For notational simplicity, hereinafter the average number of associated users of the tagged and non-tagged IAB nodes are denoted as U_s and \bar{U}_s , respectively. Similarly, each tagged IAB donor node has $N_x^m = 1 + 1.28 \frac{\lambda_s}{\lambda_m}$ associated IAB nodes, while each non-tagged IAB donor node has $N_x^m = \frac{\lambda_s}{\lambda_m}$ associated IAB nodes. The average number of associated IAB nodes of the tagged and non-tagged IAB nodes are denoted as U_m and \bar{U}_m , respectively.

III. PROPAGATION MODEL AND RATE CHARACTERIZATION

A. BLOCKAGE MODEL

Unlike FR1 band signals, FR2 signals are susceptible to blockages (concrete walls, trees, etc.) which result in both LOS and NLOS transmissions. Therefore, we consider a two-state stationary probabilistic “exponential blockage model” that is validated in [30], where the probabilities of the occurrence of the LOS and NLOS transmissions are defined by $p_{\mathcal{L}}(r) = e^{-\beta r}$ and $p_{\mathcal{N}}(r) = 1 - e^{-\beta r}$, respectively, with β and r denoting the blockage density and the link length. Accordingly, considering the blockage effect, the path loss exponent on each link is a random variable given as

$$\alpha = \begin{cases} \alpha_{\mathcal{L}} & \text{w.p. } p_{\mathcal{L}}(r) \\ \alpha_{\mathcal{N}} & \text{w.p. } p_{\mathcal{N}}(r), \end{cases} \quad (1)$$

where $\alpha_{\mathcal{L}}$ and $\alpha_{\mathcal{N}}$ are the LOS and NLOS path loss exponents, respectively.

B. FR2 CHANNEL MODEL

This work considers wide-band FR2 channel model and OFDM transmission, where the impact of the multi-path delay spread is more significant in NLOS environments. Further, each path is modeled by a time delay d , in addition to the complex gain as well as the angles of departure and arrival [31]. Therefore, the delay- d MIMO channel matrix between the transmitter at x and the receiver at y based on the geometric channel model that incorporates wide-band channel characteristics, limited FR2 scattering properties, and multi-path delays is given as

$$\mathbf{H}_{xy}[d] = \sqrt{\frac{n_t n_r}{r_{xy}^{\alpha} \eta_{xy}}} \sum_{l=1}^{\eta_{xy}} h_{lxy} p_{\text{rc}}(dT_s - \tau_l) \mathbf{a}_R(\theta_{lxy}) \mathbf{a}_T^H(\phi_{lxy}), \quad (2)$$

where $n_r \in \{n_r^m, n_r^s, n_r^u\}$ and $n_t \in \{n_t^m, n_t^s\}$ depending on the receiver and transmitter. h_{lxy} is the small-scale Rayleigh fading on the l th path [5], [32]. $\alpha \in \{\alpha_{\mathcal{L}}, \alpha_{\mathcal{N}}\}$ and $\eta_{xy} \in \{\eta_{\mathcal{L}}, \eta_{\mathcal{N}}\}$ are the path loss exponents and the number of scatters depending on LOS or NLOS path such that $\eta_{\mathcal{L}} < \eta_{\mathcal{N}}$. $\mathbf{a}_R(\theta)$ and $\mathbf{a}_T(\phi)$ are steering vectors of receiver (*i.e.*, users for $R = u$ or IAB nodes for $R = s$) and transmitter (*i.e.*, IAB nodes for $T = s$ or IAB donor nodes for $T = m$), respectively.

In particular, the steering vectors of the ULAs are respectively given as $\mathbf{a}_R(\theta) = \frac{1}{n_r} [1 e^{jk\tilde{d}\sin(\theta)} \dots e^{(n_r-1)jk\tilde{d}\sin(\theta)}]$ and $\mathbf{a}_T(\phi) = \frac{1}{n_t} [1 e^{jk\tilde{d}\sin(\phi)} \dots e^{(n_t-1)jk\tilde{d}\sin(\phi)}]$, where θ and ϕ are the angles of arrival (AOA) and the angles of departure (AOD), respectively. Further, $\tilde{k} = \frac{2\pi}{\lambda_c}$ with λ_c being the wavelength and \tilde{d} is the distance between antenna elements. $p_{rc}(\tau)$ denotes a pulse-shaping function for T_s -spaced signalling evaluated at τ seconds, where $p_{rc}(t)$ is given as

$$p_{rc}(t) = \begin{cases} \frac{\pi}{4} \text{sinc}\left(\frac{1}{2\epsilon}\right), & t = \pm \frac{T_s}{2\epsilon} \\ \text{sinc}\left(\frac{t}{T_s}\right) \frac{\cos\left(\frac{\pi\epsilon t}{T_s}\right)}{1 - \left(\frac{2\epsilon t}{T_s}\right)^2}, & \text{otherwise,} \end{cases} \quad (3)$$

where T_s is the sampling time and $\epsilon = 1$ is the roll-off factor. Given the delay- d MIMO channel matrix, the channel at subcarrier k , $\mathbf{H}_{xy}[k]$, can be characterized with the assumption of perfect synchronization as

$$\mathbf{H}_{xy}[k] = \sum_{d=0}^{D-1} \mathbf{H}_{xy}[d] e^{-j\frac{2\pi k}{K}d}, \quad (4)$$

where K is the total subcarriers in the MIMO-OFDM hybrid beamforming transmission model. D is the cyclic prefix added to the symbol blocks before applying precoding for eliminating the inter symbol interference (ISI) effect. For notational simplicity, the channel matrix at k th subcarrier is given as

$$\mathbf{H}_{xy}[k] = \sqrt{\frac{n_t n_r}{r_{xy}^{\alpha} \eta_{xy}}} \sum_{l=1}^{\eta_{xy}} h_{lxy} \omega_{\tau_l}[k] \mathbf{a}_R(\theta_{lxy}) \mathbf{a}_T^H(\phi_{lxy}), \quad (5)$$

where $\omega_{\tau_l}[k] = \sum_{d=0}^{D-1} p_{rc}(dT_s - \tau_l) e^{-j\frac{2\pi k}{K}d}$. In particular, the path delay is uniformly distributed in $[0, D T_s]$.

C. ACCESS LINK [IAB NODE-USER link]

1) RECEIVED SIGNAL

The users that are associated with the IAB node are simultaneously served through a fully-connected hybrid beamforming architecture. Let the BB and RF precoder matrices of the tagged IAB nodes located at x_s^j with $j \in \{\mathcal{L}, \mathcal{N}\}$ with respect to the typical user (located at origin) be denoted as $\mathbf{V}_{x_s^j}^{\text{BB}} = [\mathbf{v}_{x_s^j,0}^{\text{BB}}, \mathbf{v}_{x_s^j,1}^{\text{BB}}, \dots, \mathbf{v}_{x_s^j,U_s-1}^{\text{BB}}]$ and $\mathbf{V}_{x_s^j}^{\text{RF}} = [\mathbf{v}_{x_s^j,0}^{\text{RF}}, \mathbf{v}_{x_s^j,1}^{\text{RF}}, \dots, \mathbf{v}_{x_s^j,U_s-1}^{\text{RF}}]$, respectively. The BB and RF precoders for IAB donor nodes are also defined similarly. Therefore, the received signal through the k th subcarrier at the typical user from the IAB node at x_s^j after passing through the BB and RF precoders and RF combiner at the typical user is given in (6), as shown at the bottom of next page, where $\mathbf{w}_{0,x_s^j}^{\text{RF}}$ is the RF combiner of the typical user and $n_0 \sim CN(0, \sigma^2)$ denotes the noise power. Since each subcarrier is assumed to be equally allocated with the transmit power $\frac{P_{(*)}}{K}$, $s_{(*)}$ is the transmitted symbol from the IAB nodes with the average power $\frac{P_{(*)}}{UK}$, such that $U \in \{U_s, \bar{U}_s\}$ and the transmit symbol $\hat{s}_{(*)}$ from IAB donor nodes is allocated with power $\frac{P_m}{KU}$, such

that $U \in \{U_m, \bar{U}_m\}$. Now, in order to eliminate the intra-cell interference, we consider zero forcing (ZF) at BB level, and therefore we have $\mathbf{v}_{x_s^j,0}^{\text{BB}} = \bar{\mathbf{h}}_{x_s^j,0}^H (\bar{\mathbf{h}}_{x_s^j,0} (\bar{\mathbf{h}}_{x_s^j,0}^H))^{-1}$, where $\bar{\mathbf{h}}_{x_s^j,0} = (\mathbf{w}_{0,x_s^j}^{\text{RF}})^H \mathbf{H}_{x_s^j,0}[k] \mathbf{V}_{x_s^j}^{\text{RF}}$.

Remark 1: Since the channel is the function of the subcarrier k , the SINR and the sum-rate over all subcarriers is the function of k . However, k only appears in the complex exponential function, which is less than or equal to 1. Thus, to consider the best-case scenario for all k , we let $\omega_{\tau_l} \leq \sum_{d=0}^{D-1} p_{rc}(dT_s - \tau_l)$, which is independent of k . Further, the analysis of this work is conducted from a probabilistic perspective. Therefore, we consider that each path has the same path delay $\tau_l = \tau = \frac{DT_s}{2} \forall l \in \{1, \dots, \eta_{xy}\}$ from an average sense, by which we can simplify and avoid taking average with respect to delay in each path in the analysis. Thus, we can assume $\omega_{\tau_l} = \omega = \sum_{d=0}^{D-1} p_{rc}(dT_s - \tau)$ to be a constant.

Remark 2: We consider the sub-optimal approach given in [27] to obtain the hybrid precoders, whereby $\mathbf{w}_{0,x_s^j}^{\text{RF}} = \mathbf{a}_u(\theta_{l_{\max},x_s^j,0})$ and $\mathbf{v}_{x_s^j,0}^{\text{RF}} = \mathbf{a}_s(\phi_{l_{\max},x_s^j,0})$. $\theta_{l_{\max},x_s^j,0}$ and $\phi_{l_{\max},x_s^j,0}$ are chosen such that the maximum channel gain is achieved on the l_{\max} path, i.e., $l_{\max} = \arg \max_l h_{l,x_s^j,0}$. Hereinafter, we ignore the subscript l_{\max} for notational simplicity.

2) SINR CHARACTERIZATION

Based on the received signal shown in (6), the SINR of the typical user from the IAB node at x_s^j with $j \in \{\mathcal{L}, \mathcal{N}\}$ is formulated as shown in (7), at the bottom of the next page,

where $I_{s_1}^a = \sum_{\substack{v \in \mathcal{U}_j \\ x_s^v \\ v \neq 0}} \frac{P_s}{KU_s} |(\mathbf{w}_{0,x_s^j}^{\text{RF}})^H \mathbf{H}_{x_s^j,0}[k] \mathbf{V}_{x_s^v}^{\text{RF}} \mathbf{v}_{x_s^v,0}^{\text{BB}}|^2$ tends to

zero due to the application of ZF precoder [27]. $I_{s_2}^a$ and I_m^a are the interference from other IAB nodes and IAB donor nodes, respectively. In particular, for tractability, the SINR is reduced based on Remark 2 and the following assumptions [33]:

- 1) n_t^s and n_r^u are sufficiently large and
- 2) $n_t^s \gg U_s$.

For the l_{\max} path, the denominator is approximated as (8), shown at the bottom of the 8th page, where the normalized array gain is defined as

$$\begin{aligned} G(\theta_1 - \theta_2) &= |\mathbf{a}^H(\theta_1) \mathbf{a}(\theta_2)|^2 \\ &= \frac{1}{n^2} \left| \sum_{i=0}^{n-1} e^{j2\pi i(\Theta_1 - \Theta_2)} \right|^2 = \frac{\sin^2[\pi n(\Theta_1 - \Theta_2)]}{n^2 \sin^2[\pi(\Theta_1 - \Theta_2)]}, \quad (9) \end{aligned}$$

where $\Theta_1 = \sin(\theta_1)$ and $\Theta_2 = \sin(\theta_2)$. When $\theta_1 = \theta_2$, $G(\theta_1 - \theta_2) = 1$; otherwise, it is approximated to zero due to the ON/OFF model [27]. Therefore, by using the above results, the SINR can be reduced to

$$\text{SINR}_{x_s^j}[k] \approx \frac{\frac{P_s}{KU_s} \frac{n_t^s n_r^u}{\eta_j} |h_{x_s^j,0}|^2 |\omega|^2 r_{x_s^j,0}^{-\alpha_j} P_{\text{ZF}}}{I_{s_2}^a + I_m^a + \sigma^2}, \quad (10)$$

where p_{ZF} is the ZF precoding penalty⁴ defined as

$$p_{ZF} = \begin{cases} 1 & \text{w.p. } (1 - \frac{1}{n_t^s})^{(U_s-1)} \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Besides, based on [27], the interference can be further simplified by neglecting the effect of ZF on interference links and when the number of antennas is large. Accordingly, we have the approximated interference as

$$I_{s2}^a = \sum_{\hat{j} \in \{\mathcal{L}, \mathcal{N}\}} \sum_{\substack{b \in \Phi_s^{\hat{j}}, \\ b \neq x_s^{\hat{j}}}} \frac{P_s}{K \bar{U}_s} \frac{n_t^s n_r^u}{\eta_j} r_{b0}^{-\alpha_j} \\ \times \sum_{u \in \mathcal{U}_b} \left| \sum_{\hat{l}=1}^{\eta_j} h_{l b0} \omega \underbrace{\mathbf{a}_u^H(\theta_{x_s^{\hat{j}}}) \mathbf{a}_u(\theta_{l b0}) \mathbf{a}_s^H(\phi_{l b0}) \mathbf{a}_s(\phi_{bu})}_{\gamma_{las}} \right|^2. \quad (12)$$

Nevertheless, we consider a modified ON/OFF model since the ON/OFF model underestimates the interference that may not be always accurate [27]. Accordingly, we have

$$\gamma_{las} = \begin{cases} 1 & \text{if } \theta_{x_s^{\hat{j}}} = \theta_{l b0}, \phi_{l b0} = \phi_{bu} \\ \rho_{BS} & \text{if } \theta_{x_s^{\hat{j}}} \neq \theta_{l b0}, \phi_{l b0} = \phi_{bu} \\ \rho_{UE} & \text{if } \theta_{x_s^{\hat{j}}} = \theta_{l b0}, \phi_{l b0} \neq \phi_{bu} \\ \rho_{BS} \rho_{UE} & \text{otherwise,} \end{cases} \quad (13)$$

where $\rho_{BS} < 1$ and $\rho_{UE} < 1$. In the subsequent section, we will use (13) to simplify the distribution of interference.

In a similar way, we have

$$I_m^a = \sum_{\tilde{j} \in \{\mathcal{L}, \mathcal{N}\}} \sum_{t \in \Phi_m^{\tilde{j}}} \frac{P_m}{K \bar{U}_m} \frac{n_t^m n_r^u}{r_{t0}^{\alpha_j} \eta_j} \\ \times \sum_{q \in \mathcal{U}_t} \left| \sum_{\hat{l}=1}^{\eta_j} h_{l t0} \omega \underbrace{\mathbf{a}_u^H(\theta_{x_s^{\tilde{j}}}) \mathbf{a}_u(\theta_{l t0}) \mathbf{a}_m^H(\phi_{l t0}) \mathbf{a}_m(\phi_{lq})}_{\gamma_{lam}} \right|^2, \quad (14)$$

where γ_{lam} is similar to (13).

⁴For more details on the ZF penalty, the readers are directed to [33].

D. BACKHAUL LINK [IAB DONOR NODE-IAB NODE link]

1) RECEIVED SIGNAL

This work considers the typical backhaul link [5], where the receiver (the tagged IAB node) served by the serving IAB donor node located at x_m^j with $j \in \{\mathcal{L}, \mathcal{N}\}$ is considered to be located at the origin, denoted as $\hat{0}$. Similarly, after passing through the analog combiner of the tagged IAB node, denoted as $\mathbf{w}_{\hat{0}x_m^j}^{\text{RF}}$, and the wideband geometric channel model $\mathbf{H}_{x_m^j \hat{0}}$ given in (2), and the BB and RF precoders of the IAB donor node, the received signal⁵ of the k th subcarrier at the tagged IAB node is expressed as (15), shown at the bottom of the next page, where RSI is the remaining residual self-interference after necessary analog/digital interference cancellation scheme has been applied. In particular, the RSI is mainly contributed by the transmit power. Accordingly, to capture the effects of IBFD operation in the IAB network the remaining RSI is modelled as $\eta_{\text{RSI}} \frac{P_s}{K}$ for each subcarrier, where η_{RSI} is the tuning parameter (RSIC factor) controlling the amount of RSI remaining in the system, which depends on the efficiency of RSIC [3], [26], [34]. Other parameters are similar to those defined for the access link.

2) SINR CHARACTERIZATION

Based on (15), we can write the SINR as (16), shown at the bottom of the 9th page. Here, the intra cell interference $I_{m1}^b = \sum_{\substack{v \in \mathcal{U}_{x_m^j}, \\ v \neq \hat{0}}} \frac{P_m}{U_m K} \left| (\mathbf{w}_{\hat{0}x_m^j}^{\text{RF}})^H \mathbf{H}_{x_m^j \hat{0}} [k] \mathbf{V}_{x_m^j}^{\text{RF}} \mathbf{v}_{x_m^j}^{\text{BB}} \right|^2$ tends to zero due to the application of ZF precoder.

Based on the aforementioned discussion, the simplified SINR is given as

$$\text{SINR}_{x_m^j} [k] = \frac{\frac{P_m}{U_m K} \frac{n_t^m n_r^s}{\eta_j} r_{x_m^j \hat{0}}^{-\alpha_j} |h_{x_m^j \hat{0}}|^2 |\omega|^2 p_{ZF}}{I_{m2}^b + I_s^b + \eta_{\text{RSI}} \frac{P_s}{K} + \sigma^2}, \quad (17)$$

⁵We consider all transmitted signals towards the cache-miss users to be encoded together as one symbol.

$$y_{0x_s^j} [k] = (\mathbf{w}_{0x_s^j}^{\text{RF}})^H \mathbf{H}_{x_s^j 0} [k] \mathbf{V}_{x_s^j}^{\text{RF}} \mathbf{v}_{x_s^j}^{\text{BB}} s_{0x_s^j} + \sum_{\substack{v \in \mathcal{U}_{x_s^j}, \\ v \neq \hat{0}}} (\mathbf{w}_{vx_s^j}^{\text{RF}})^H \mathbf{H}_{x_s^j v} [k] \mathbf{V}_{x_s^j}^{\text{RF}} \mathbf{v}_{x_s^j}^{\text{BB}} s_{vx_s^j} \\ + \sum_{\substack{b \in \Phi_s, u \in \mathcal{U}_b, \\ b \neq x_s^j}} (\mathbf{w}_{0x_s^j}^{\text{RF}})^H \mathbf{H}_{b0} [k] \mathbf{V}_b^{\text{RF}} \mathbf{v}_{bu}^{\text{BB}} s_{ub} + \sum_{t \in \Phi_m} \sum_{q \in \mathcal{U}_t} (\mathbf{w}_{0x_s^j}^{\text{RF}})^H \mathbf{H}_{t0} [k] \mathbf{V}_t^{\text{RF}} \mathbf{v}_{tq}^{\text{BB}} \hat{s}_{qt} + n_0 \quad (6)$$

$$\text{SINR}_{x_s^j} [k] = \frac{\frac{P_s}{U_s K} \left| (\mathbf{w}_{0x_s^j}^{\text{RF}})^H \mathbf{H}_{x_s^j 0} [k] \mathbf{V}_{x_s^j}^{\text{RF}} \mathbf{v}_{x_s^j}^{\text{BB}} \right|^2}{I_{s1}^a + \underbrace{\sum_{\substack{b \in \Phi_s, u \in \mathcal{U}_b, \\ b \neq x_s^j}} \frac{P_s}{K \bar{U}_s} \left| (\mathbf{w}_{0x_s^j}^{\text{RF}})^H \mathbf{H}_{b0} [k] \mathbf{V}_b^{\text{RF}} \mathbf{v}_{bu}^{\text{BB}} \right|^2}_{I_{s2}^a} + \underbrace{\sum_{t \in \Phi_m} \sum_{q \in \mathcal{U}_t} \frac{P_m}{K \bar{U}_m} \left| (\mathbf{w}_{0x_s^j}^{\text{RF}})^H \mathbf{H}_{t0} [k] \mathbf{V}_t^{\text{RF}} \mathbf{v}_{tq}^{\text{BB}} \right|^2}_{I_m^a} + \sigma^2} \quad (7)$$

where the approximated interference $I_{m_2}^b$ and I_s^b are respectively given as

$$I_{m_2}^b = \sum_{\hat{j} \in \{\mathcal{L}, \mathcal{N}\}} \frac{P_m}{K \bar{U}_m} \sum_{\substack{t \in \Phi_m^{\hat{j}}, \\ t \neq x_m}} \frac{n_t^m n_r^s}{r_{t\hat{0}}^{\alpha_j} \eta_j} \times \sum_{u \in \mathcal{L}_t} \left| \sum_{\hat{l}=1}^{\eta_j} h_{l\hat{t}0} \omega \underbrace{\mathbf{a}_u^H(\theta_{x_m \hat{0}}) \mathbf{a}_u(\theta_{l\hat{t}\hat{0}}) \mathbf{a}_m^H(\phi_{l\hat{t}\hat{0}}) \mathbf{a}_m(\phi_u)}_{\gamma_{l\hat{t}m} \text{ that is similar to (13)}} \right|^2, \quad (18)$$

and

$$I_s^b = \sum_{\hat{j} \in \{\mathcal{L}, \mathcal{N}\}} \frac{P_s}{K \bar{U}_s} \sum_{b \in \Phi_s^{\hat{j}}} \frac{n_t^s n_r^s}{r_{b\hat{0}}^{\alpha_j} \eta_j} \times \sum_{q \in \mathcal{L}_t} \left| \sum_{\hat{l}=1}^{\eta_j} h_{l\hat{t}0} \omega \underbrace{\mathbf{a}_u^H(\theta_{x_m \hat{0}}) \mathbf{a}_u(\theta_{l\hat{t}\hat{0}}) \mathbf{a}_m^H(\phi_{l\hat{t}\hat{0}}) \mathbf{a}_m(\phi_{bq})}_{\gamma_{l\hat{t}m} \text{ that is similar to (13)}} \right|^2. \quad (19)$$

Due to aforementioned, the rates of access and backhaul over all subcarriers are given as

$$R_a^j = \sum_{k=1}^K \frac{W}{K} \log(1 + \text{SINR}_{x_s^j}[k]), \quad (20)$$

$$R_b^j = \sum_{k=1}^K \frac{W}{K} \log(1 + \text{SINR}_{x_m^j}[k]). \quad (21)$$

Further, based on Remark 1, the rates are simplified as

$$R_a^j = W \log(1 + \text{SINR}_{x_s^j}), \quad (22)$$

$$R_b^j = W \log(1 + \text{SINR}_{x_m^j}), \quad (23)$$

where $j \in \{\mathcal{L}, \mathcal{N}\}$.

IV. PERFORMANCE ANALYSIS

In this section we consider three key performance metrics, namely ASP of file delivery, latency, and throughput for

file delivery, that we will use to evaluate the performance of the considered system model. In order to derive these metrics, we define the path loss process through mapping theorem to give its intensity measure and density so that the association probabilities where the typical user is associated with the IAB node and the typical IAB node is associated with the IAB donor node in LOS and NLOS path are given, respectively.

A. ASSOCIATION PROBABILITIES

We define the sequence $\mathcal{L}_t = \{||t||^\alpha : t \in \Phi_i\} = \{||t||^{\alpha_{\mathcal{L}}} : t^{\mathcal{L}} \in \Phi_i^{\mathcal{L}}\} \cup \{||t||^{\alpha_{\mathcal{N}}} : t^{\mathcal{N}} \in \Phi_i^{\mathcal{N}}\}$ as the path loss process linked to Φ_i with $i \in \{s, m\}$, where t is the location of the associated BS ($t = x_s$ for $i = s$ and $t = x_m$ for $i = m$).

Lemma 1: The intensity measure of the path loss process \mathcal{L}_t is given by

$$\Lambda_{\mathcal{L}_t}([0, l]) = 2\pi \lambda_i \left(Z \left(l^{\frac{1}{\alpha_{\mathcal{L}}}} \right) + \hat{Z} \left(l^{\frac{1}{\alpha_{\mathcal{N}}}} \right) \right), \quad (24)$$

where $Z(l) = -\frac{1}{\beta} l e^{-\beta l} - \frac{1}{\beta^2} (e^{-\beta l} - 1)$ and $\hat{Z}(l) = \frac{l^2}{2} - Z(l)$ and $i \in \{s, m\}$. And the density is given as $\lambda_{\mathcal{L}_t}(l) =$

$$2\pi \lambda_i \left(Z' \left(l^{\frac{1}{\alpha_{\mathcal{L}}}} \right) \frac{1}{\alpha_{\mathcal{L}}} l^{\frac{1}{\alpha_{\mathcal{L}}}-1} + \hat{Z}' \left(l^{\frac{1}{\alpha_{\mathcal{N}}}} \right) \frac{1}{\alpha_{\mathcal{N}}} l^{\frac{1}{\alpha_{\mathcal{N}}}-1} \right), \quad (25)$$

where $Z'(l) = e^{-\beta l} l$ and $\hat{Z}'(l) = l - e^{-\beta l}$ and $i \in \{s, m\}$.

Proof: The proof is given in Appendix A. \square

In particular, based on the proof, $\Lambda_{\mathcal{L}_{x_i}}([0, l]) = \Lambda_{\mathcal{L}_{x_i^{\mathcal{L}}}}([0, l]) + \Lambda_{\mathcal{L}_{x_i^{\mathcal{N}}}}([0, l])$, where $\Lambda_{\mathcal{L}_{x_i^{\mathcal{L}}}}([0, l]) = 2\pi \lambda_i Z(l^{\frac{1}{\alpha_{\mathcal{L}}}})$ and $\Lambda_{\mathcal{L}_{x_i^{\mathcal{N}}}}([0, l]) = 2\pi \lambda_i \hat{Z}(l^{\frac{1}{\alpha_{\mathcal{N}}}})$ with $i \in \{s, m\}$. Similarly, $\lambda_{\mathcal{L}_{x_i^{\mathcal{L}}}}(l) = 2\pi \lambda_i Z'(l^{\frac{1}{\alpha_{\mathcal{L}}}}) \frac{1}{\alpha_{\mathcal{L}}} l^{\frac{1}{\alpha_{\mathcal{L}}}-1}$ and $\lambda_{\mathcal{L}_{x_i^{\mathcal{N}}}}(l) = 2\pi \lambda_i \hat{Z}'(l^{\frac{1}{\alpha_{\mathcal{N}}}}) \frac{1}{\alpha_{\mathcal{N}}} l^{\frac{1}{\alpha_{\mathcal{N}}}-1}$ are acquired by taking derivative of the intensity measure of the sub path loss process in terms of l .

Having derived the intensity measure and density, the PDF of the ‘‘contact distance’’ L_t is derived by taking derivative with respect to the void probability that is given as $\mathbb{P}[L_t \leq l] = 1 - e^{-\Lambda_{\mathcal{L}_t}([0, l])}$. The PDF of L_t is given as

$$f_{L_t}(l) = \lambda_{\mathcal{L}_t}(l) \exp[-\Lambda_{\mathcal{L}_t}([0, l])], \quad l \geq 0, \quad (26)$$

$$\begin{aligned} \frac{P_s}{KU_s} \left| (\mathbf{w}_{0x_s^j}^{\text{RF}})^H \mathbf{H}_{x_s^j 0}[k] \mathbf{V}_{x_s^j}^{\text{RF}} \mathbf{v}_{x_s^j 0}^{\text{BB}} \right|^2 &\approx \frac{P_s}{KU_s} \left| \mathbf{a}_u^H(\theta_{x_s^j 0}) \sqrt{\frac{n_t^s n_r^u}{r_{x_s^j 0}^{\alpha_j} \eta_j}} h_{x_s^j 0} \omega \mathbf{a}_u(\theta_{x_s^j 0}) \mathbf{a}_s(\phi_{x_s^j 0})^H \mathbf{a}_s(\phi_{x_s^j 0}) \right|^2 PZF \\ &= \frac{P_s}{KU_s} \frac{n_t^s n_r^u}{r_{x_s^j 0}^{\alpha_j} \eta_j} |h_{x_s^j 0}|^2 |\omega|^2 \left| \mathbf{a}_u^H(\theta_{x_s^j 0}) \mathbf{a}_u(\theta_{x_s^j 0}) \mathbf{a}_s(\phi_{x_s^j 0})^H \mathbf{a}_s(\phi_{x_s^j 0}) \right|^2 PZF \end{aligned} \quad (8)$$

$$\begin{aligned} y_{\hat{0}x_m^j}[k] &= (\mathbf{w}_{\hat{0}x_m^j}^{\text{RF}})^H \mathbf{H}_{x_m^j \hat{0}}[k] \mathbf{V}_{x_m^j}^{\text{RF}} \mathbf{v}_{x_m^j \hat{0}}^{\text{BB}} \hat{s}_{\hat{0}x_m^j} + \sum_{\substack{v \in \Phi_m \\ v \neq \hat{0}}} (\mathbf{w}_{\hat{0}x_m^j}^{\text{RF}})^H \mathbf{H}_{x_m^j v}[k] \mathbf{V}_{x_m^j}^{\text{RF}} \mathbf{v}_{x_m^j v}^{\text{BB}} \hat{s}_{v x_m^j} \\ &+ \sum_{\substack{t \in \Phi_m \\ t \neq x_m^j}} \sum_{u \in \mathcal{L}_t} (\mathbf{w}_{\hat{0}x_m^j}^{\text{RF}})^H \mathbf{H}_{t \hat{0}}[k] \mathbf{V}_t^{\text{RF}} \mathbf{v}_{tu}^{\text{BB}} \hat{s}_{\hat{0}t} + \sum_{b \in \Phi_s} \sum_{q \in \mathcal{L}_b} (\mathbf{w}_{\hat{0}x_m^j}^{\text{RF}})^H \mathbf{H}_{b \hat{0}}[k] \mathbf{V}_b^{\text{RF}} \mathbf{v}_{bq}^{\text{BB}} \hat{s}_{ub} + \text{RSI} + n_{\hat{0}} \end{aligned} \quad (15)$$

where $t \in \{x_s, x_m\}$. Similarly, based on the intensity measure of sub path loss process with respect to LOS and NLOS paths, $f_{L_{x_i^{\mathcal{L}}}}(l) = \lambda_{\mathcal{L}_{x_i^{\mathcal{L}}}}(l) \exp[-\Lambda_{\mathcal{L}_{x_i^{\mathcal{L}}}}([0, l])]$ and $f_{L_{x_i^{\mathcal{N}}}}(l) = \lambda_{\mathcal{L}_{x_i^{\mathcal{N}}}} \exp[-\Lambda_{\mathcal{L}_{x_i^{\mathcal{N}}}}([0, l])]$, where the $\lambda_{\mathcal{L}_{x_i^{\mathcal{L}}}}(l)$ and $\lambda_{\mathcal{L}_{x_i^{\mathcal{N}}}}(l)$ are given above, with $i \in \{\mathcal{L}, \mathcal{N}\}$.

Since the channels of access and backhaul are composed of multiple paths (either LOS or NLOS), we give the association probabilities such that the typical user (IAB node) is associated with the IAB node (IAB donor node) with LOS and NLOS channels below.

Proposition 1: The association probabilities that the typical user is associated with the IAB node with either LOS or NLOS channel, denoted as $x_s^{\mathcal{L}}$ and $x_s^{\mathcal{N}}$, are given as

$$\mathcal{A}_s^{\mathcal{L}} = \int_0^\infty \exp\left[-2\pi\lambda_s\hat{Z}\left(l^{\frac{1}{\alpha_{\mathcal{N}}}}\right) - 2\pi\lambda_s Z\left(l^{\frac{1}{\alpha_{\mathcal{L}}}}\right)\right] \lambda_{\mathcal{L}_{x_s^{\mathcal{L}}}}(l) dl, \quad (27)$$

$$\mathcal{A}_s^{\mathcal{N}} = \int_0^\infty \exp\left[-2\pi\lambda_s Z\left(l^{\frac{1}{\alpha_{\mathcal{L}}}}\right) - 2\pi\lambda_s\hat{Z}\left(l^{\frac{1}{\alpha_{\mathcal{N}}}}\right)\right] \lambda_{\mathcal{L}_{x_s^{\mathcal{N}}}}(l) dl. \quad (28)$$

Similarly, the association probabilities that the typical IAB node is associated with the IAB donor node with either LOS or NLOS channel, denoted as $x_m^{\mathcal{L}}$ and $x_m^{\mathcal{N}}$, are given as

$$\mathcal{A}_m^{\mathcal{L}} = \int_0^\infty \exp\left[-2\pi\lambda_m\hat{Z}\left(l^{\frac{1}{\alpha_{\mathcal{N}}}}\right) - 2\pi\lambda_s Z\left(l^{\frac{1}{\alpha_{\mathcal{L}}}}\right)\right] \lambda_{\mathcal{L}_{x_m^{\mathcal{L}}}}(l) dl, \quad (29)$$

$$\mathcal{A}_m^{\mathcal{N}} = \int_0^\infty \exp\left[-2\pi\lambda_m Z\left(l^{\frac{1}{\alpha_{\mathcal{L}}}}\right) - 2\pi\lambda_s\hat{Z}\left(l^{\frac{1}{\alpha_{\mathcal{N}}}}\right)\right] \lambda_{\mathcal{L}_{x_m^{\mathcal{N}}}}(l) dl, \quad (30)$$

where $\lambda_{\mathcal{L}_{x_s^{\mathcal{L}}}}(l)$, $\lambda_{\mathcal{L}_{x_s^{\mathcal{N}}}}(l)$, $\lambda_{\mathcal{L}_{x_m^{\mathcal{L}}}}(l)$, and $\lambda_{\mathcal{L}_{x_m^{\mathcal{N}}}}(l)$ are given above.

Proof: The proof is given in Appendix B. \square

Having derived the association probabilities, we now have the probability density functions (PDFs) of the ‘‘serving distances’’ in LOS and NLOS between the serving IAB node and the typical user, which are given as [35]

$$\hat{f}_{L_{x_s^{\mathcal{L}}}}(l) = \frac{1}{\mathcal{A}_{\mathcal{L}}} \exp\left[-\Lambda_{\mathcal{L}_{x_s^{\mathcal{L}}}}([0, l])\right] f_{L_{x_s^{\mathcal{L}}}}(l), \quad (31)$$

$$\hat{f}_{L_{x_s^{\mathcal{N}}}}(l) = \frac{1}{\mathcal{A}_{\mathcal{N}}} \exp\left[-\Lambda_{\mathcal{L}_{x_s^{\mathcal{N}}}}([0, l])\right] f_{L_{x_s^{\mathcal{N}}}}(l). \quad (32)$$

Similarly, the PDF of the ‘‘serving distances’’ between the serving IAB donor node and the typical IAB node are given

as

$$\hat{f}_{L_{x_m^{\mathcal{L}}}}(l) = \frac{1}{\mathcal{A}_{\mathcal{L}}} \exp\left[-\Lambda_{\mathcal{L}_{x_m^{\mathcal{L}}}}([0, l])\right] f_{L_{x_m^{\mathcal{L}}}}(l), \quad (33)$$

$$\hat{f}_{L_{x_m^{\mathcal{N}}}}(l) = \frac{1}{\mathcal{A}_{\mathcal{N}}} \exp\left[-\Lambda_{\mathcal{L}_{x_m^{\mathcal{N}}}}([0, l])\right] f_{L_{x_m^{\mathcal{N}}}}(l). \quad (34)$$

B. ASP OF FILE DELIVERY

The ASP of file delivery is defined as the probability of rate coverage where the instantaneous rate is greater than a pre-defined threshold rate. With the association probabilities, the ASP of file delivery is computed by the summation of each conditional ASP of file delivery. Thus, the ASP of file delivery under IBFD mode for the IAB network is given as

$$\begin{aligned} \mathcal{P}_s^{\text{IBFD}}(\nu) &= \sum_{i=1}^F q_i \left\{ \mathcal{A}_s^{\mathcal{L}} \left[\tilde{\omega}_i \mathbb{P}[\mathbf{R}_a^{\mathcal{L}} \geq \nu] + (1 - \tilde{\omega}_i) \mathbb{P}[\mathbf{R}_a^{\mathcal{L}} \geq \nu] (\mathcal{A}_m^{\mathcal{L}} \mathbb{P}[\mathbf{R}_b^{\mathcal{L}} \geq \nu] \right. \right. \\ &\quad \left. \left. + \mathcal{A}_m^{\mathcal{N}} \mathbb{P}[\mathbf{R}_b^{\mathcal{N}} \geq \nu]) \right] + \mathcal{A}_s^{\mathcal{N}} \left[\tilde{\omega}_i \mathbb{P}[\mathbf{R}_a^{\mathcal{N}} \geq \nu] \right. \right. \\ &\quad \left. \left. + (1 - \tilde{\omega}_i) \mathbb{P}[\mathbf{R}_a^{\mathcal{N}} \geq \nu] (\mathcal{A}_m^{\mathcal{L}} \mathbb{P}[\mathbf{R}_b^{\mathcal{L}} \geq \nu] + \mathcal{A}_m^{\mathcal{N}} \mathbb{P}[\mathbf{R}_b^{\mathcal{N}} \geq \nu]) \right] \right\}. \end{aligned} \quad (35)$$

Proof: The proof is given in Appendix C. \square

Similarly, the ASP of file delivery for TDD and FDD modes are computed with the following changes: 1) the RSIC factor is equal to zero, 2) the interference under FDD mode at IAB nodes and the typical user will not include the other tiers, which is different from the IBFD and TDD modes, and 3) for TDD mode, the rate in the cache hit access link will not be scaled by 1/2 while for FDD mode, the rate in all access and backhaul link are scaled by 1/2 irrespective of cache hit or miss.

C. AVERAGE THROUGHPUT OF THE TYPICAL USER

In order to evaluate the network capacity for the access and backhaul links, we define the average rate of file delivery of all links in the network for different caching strategies based on the following equations:

$$\begin{aligned} \bar{R} &= \sum_{i=1}^F q_i \left\{ \mathcal{A}_s^{\mathcal{L}} \left[\tilde{\omega}_i \bar{R}_a^{\mathcal{L}} + (1 - \tilde{\omega}_i) \min(\bar{R}_a^{\mathcal{L}}, \mathcal{A}_m^{\mathcal{L}} \bar{R}_b^{\mathcal{L}} + \mathcal{A}_m^{\mathcal{N}} \bar{R}_b^{\mathcal{N}}) \right] \right. \\ &\quad \left. + \mathcal{A}_s^{\mathcal{N}} \left[\tilde{\omega}_i \bar{R}_a^{\mathcal{N}} + (1 - \tilde{\omega}_i) \min(\bar{R}_a^{\mathcal{N}}, \mathcal{A}_m^{\mathcal{L}} \bar{R}_b^{\mathcal{L}} + \mathcal{A}_m^{\mathcal{N}} \bar{R}_b^{\mathcal{N}}) \right] \right\}, \end{aligned} \quad (36)$$

$$\begin{aligned} \text{SINR}_{x_m^j}[k] &= \frac{\frac{P_m}{K\bar{U}_m} \left| (\mathbf{w}_{\hat{0}_{x_m^j}}^{\text{RF}})^H \mathbf{H}_{x_m^j \hat{0}}[k] \mathbf{V}_{x_m^j}^{\text{RF}} \mathbf{v}_{x_m^j}^{\text{BB}} \right|^2}{I_{m_1}^b + \eta_{\text{RSI}} \frac{P_s}{K} + \sigma^2}}{\underbrace{\sum_{\substack{t \in \Phi_m \\ t \neq x_m^j}} \sum_{u \in \mathcal{U}_t} \frac{P_m}{K\bar{U}_m} \left| (\mathbf{w}_{\hat{0}_{x_m^j}}^{\text{RF}})^H \mathbf{H}_{t \hat{0}}[k] \mathbf{V}_t^{\text{RF}} \mathbf{v}_{x_m^j}^{\text{BB}} \right|^2}_{I_{m_2}^b}} + \underbrace{\sum_{b \in \Phi_s} \sum_{q \in \mathcal{U}_b} \frac{P_s}{K\bar{U}_s} \left| (\mathbf{w}_{\hat{0}_{x_m^j}}^{\text{RF}})^H \mathbf{H}_{b \hat{0}}[k] \mathbf{V}_b^{\text{RF}} \mathbf{v}_{bq}^{\text{BB}} \right|^2}_{I_s^b}} \end{aligned} \quad (16)$$

where the average rates in (36) are given by [36]

$$\bar{R}_a^{\mathcal{L}} = W \int_0^\infty \mathbb{P}[\log(1 + \text{SINR}_{x_s^{\mathcal{L}}}) \geq \nu] d\nu, \quad (37)$$

$$\bar{R}_a^{\mathcal{N}} = W \int_0^\infty \mathbb{P}[\log(1 + \text{SINR}_{x_s^{\mathcal{N}}}) \geq \nu] d\nu, \quad (38)$$

$$\bar{R}_b^{\mathcal{L}} = W \int_0^\infty \mathbb{P}[\log(1 + \text{SINR}_{x_m^{\mathcal{L}}}) \geq \nu] d\nu, \quad (39)$$

$$\bar{R}_b^{\mathcal{N}} = W \int_0^\infty \mathbb{P}[\log(1 + \text{SINR}_{x_m^{\mathcal{N}}}) \geq \nu] d\nu. \quad (40)$$

The derivation of the average rates is similar to that of the ASP of file delivery. Further, ASP of file delivery and throughput for HD mode can be derived in a similar way.

D. LATENCY OF FILE DELIVERY

In order to evaluate the time required for file delivery in the network, we evaluate the latency of the network under IBFD operation. For comparison, we also calculate the same for HD mode. Under HD mode, the non-cached content will be retrieved in two hops at different time instances. However, under IBFD mode, the two hops involving retrieving and delivering can be performed at the same time.⁶ Accordingly, by a slight abuse of notations,⁷ the latency of file delivery in the network for IBFD and HD mode of operation can respectively be given as

$$T_{\text{IBFD}} = \sum_{i=1}^F q_i \left\{ \mathcal{A}_s^{\mathcal{L}} \left[\tilde{\omega}_i \frac{S}{\bar{R}_a^{\mathcal{L}}} + (1 - \tilde{\omega}_i) \max \left(\frac{S}{\bar{R}_a^{\mathcal{L}}}, \frac{S}{\bar{R}_b} \right) \right] + \mathcal{A}_s^{\mathcal{N}} \left[\tilde{\omega}_i \frac{S}{\bar{R}_a^{\mathcal{N}}} + (1 - \tilde{\omega}_i) \max \left(\frac{S}{\bar{R}_a^{\mathcal{N}}}, \frac{S}{\bar{R}_b} \right) \right] \right\}, \quad (41)$$

$$T_{\text{FDD}} = \sum_{i=1}^F q_i \left\{ \mathcal{A}_s^{\mathcal{L}} \left[\tilde{\omega}_i \frac{2S}{\bar{R}_a^{\mathcal{L}}} + (1 - \tilde{\omega}_i) \max \left(\frac{2S}{\bar{R}_a^{\mathcal{L}}}, \frac{2S}{\bar{R}_b} \right) \right] + \mathcal{A}_s^{\mathcal{N}} \left[\tilde{\omega}_i \frac{2S}{\bar{R}_a^{\mathcal{N}}} + (1 - \tilde{\omega}_i) \max \left(\frac{2S}{\bar{R}_a^{\mathcal{N}}}, \frac{2S}{\bar{R}_b} \right) \right] \right\}, \quad (42)$$

$$T_{\text{TDD}} = \sum_{i=1}^F q_i \left\{ \mathcal{A}_s^{\mathcal{L}} \left[\tilde{\omega}_i \frac{S}{\bar{R}_a^{\mathcal{L}}} + (1 - \tilde{\omega}_i) \left(\frac{S}{\frac{1}{2}\bar{R}_a^{\mathcal{L}}} + \frac{S}{\frac{1}{2}\bar{R}_b} \right) \right] + \mathcal{A}_s^{\mathcal{N}} \left[\tilde{\omega}_i \frac{S}{\bar{R}_a^{\mathcal{N}}} + (1 - \tilde{\omega}_i) \left(\frac{S}{\frac{1}{2}\bar{R}_a^{\mathcal{N}}} + \frac{S}{\frac{1}{2}\bar{R}_b} \right) \right] \right\}, \quad (43)$$

where $\bar{R}_b = \mathcal{A}_m^{\mathcal{L}} \bar{R}_b^{\mathcal{L}} + \mathcal{A}_m^{\mathcal{N}} \bar{R}_b^{\mathcal{N}}$.

V. NUMERICAL RESULTS

In this section, we numerically evaluate the performance of WEC-enabled IBFD IAB networks and compare it with i) baseline HD solutions (FDD and TDD) and ii) NC IAB scenario. Accordingly, we holistically evaluate the network

⁶Before delivering the content to the typical user in the access link, the content is retrieved through the backhaul link. Once a portion of the content is retrieved at the tagged IAB node, it can be immediately delivered to the typical user. Though a small non-simultaneous latency exists between the two phases, it is neglected for the sake of tractability.

⁷Even though the rates under FDD, TDD, and IBFD modes are different, we use the same notation to refer to the rate under all modes.

performance with respect to three key performance metrics, namely the ASP, throughput, and latency of file delivery under two popular caching strategies, MC and UC. Unless otherwise stated, the following parameters are used to produce the numerical results [17], [18], [37]: $\lambda_m = 10^{-5}$ nodes/m², $\lambda_s = 5 \times 10^{-5}$ nodes/m², $\lambda_u = 2.5 \times 10^{-4}$ nodes/m², $P_m = 46$ dBm, $P_s = 30$ dBm, $n_t^m = n_t^s = n_r^s = 256$, $n_r^u = 32$, transmit RF chain number of each IAB node $n_{\text{RF}}^m = n_{\text{RF}}^s = 10$, $\alpha_{\mathcal{L}} = 2$, $\alpha_{\mathcal{N}} = 4$, $W = 10^9$ Hz, $\nu = W \times 10^{-3}$ bits/s/Hz, $\eta_{\text{RSI}} = 10^{-8}$, $\nu = 0.8$, $\eta_{\mathcal{L}} = 3$, $\eta_{\mathcal{N}} = 5$, $\beta = 0.008$, $S = 10^6$ bits, $K = 512$, $D = 128$. Below we evaluate the trade-offs and effects of several parameters on the performance of the considered system model.

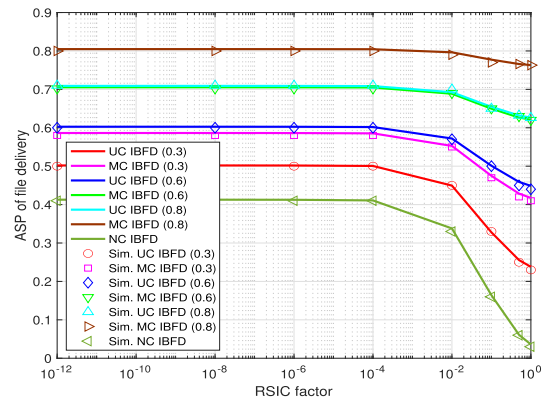


FIGURE 2. ASP of file delivery v.s. RSIC factor.

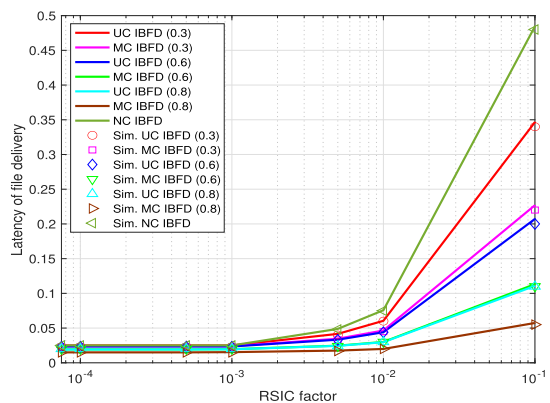


FIGURE 3. Latency of file delivery v.s. RSIC factor.

A. EFFECT OF RSIC FACTOR

We begin by evaluating the effects of RSIC on the performance of ASP, latency, and throughput of file delivery in the IAB network in Fig. 2, 3, and 4, respectively. In particular, the value inside the bracket shown in the legends of the figures signify normalized cache size. The RSIC factor denotes the level of RSI cancellation in the IBFD-enabled IAB nodes (lower values denote higher cancellation). It can be seen from the figures that the performance degrades when the RSIC factor is high. However, it can also be seen from the

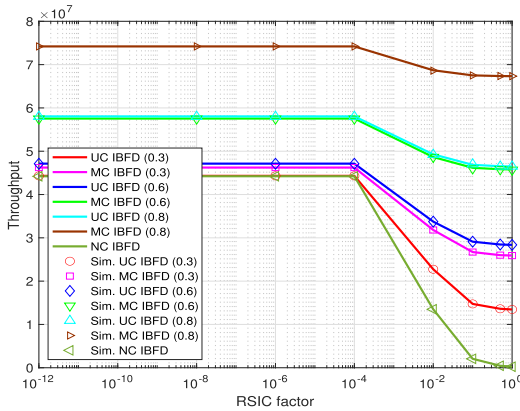


FIGURE 4. Throughput of file delivery v.s. RSIC factor.

figures that increasing the cache size reduces the impact of the RSIC factor on the performance of the IAB network. This result is not only innovative but also exciting as it provides an alternative to the technologically difficult and expensive RSIC, whereby instead of striving to achieve higher levels of RSIC through various analog and digital cancellation techniques, it may be effective to increase cache size and improve the caching probability of the files at the edge nodes to improve the performance.

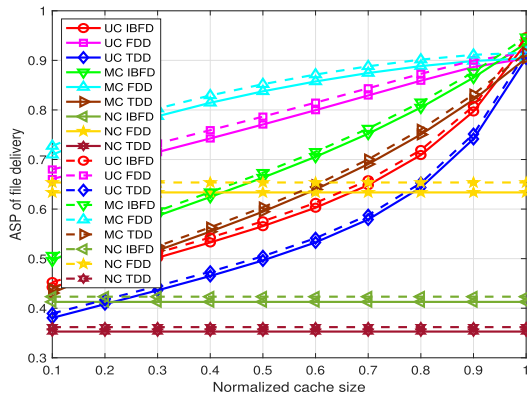


FIGURE 5. ASP of file delivery v.s. normalized cache size with different antenna numbers.

B. EFFECT OF CACHE SIZE

Next, we evaluate the ASP, latency, and throughput of file delivery with respect to the normalized cache size in Fig. 5, 6, and 7, respectively. It can be seen from the figures that when the cache size increases, more files are available in the local caches, which in turn makes the cache hit probability higher. This reduces the backhaul usage and shortens the communication distance. In particular, we can see from the figures that when the normalized cache size increases, the ASP of file delivery and the average throughput also increases, while the latency of file delivery decreases.

Moreover, in Fig. 5, the dashed lines are generated with respect to antenna numbers ($n_t^m = n_t^s = n_r^s = 512$),

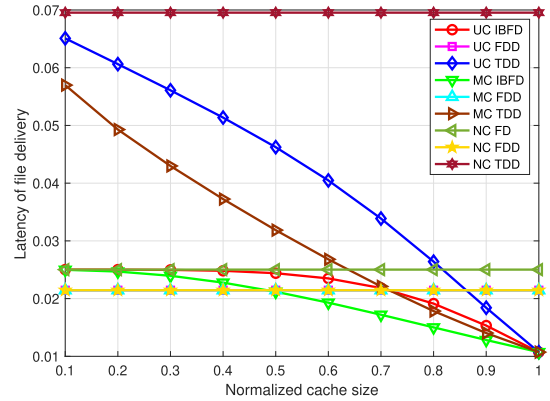


FIGURE 6. Latency of file delivery v.s. normalized cache size.

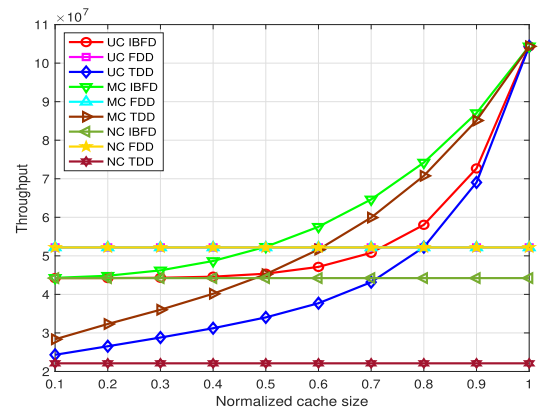


FIGURE 7. Throughput of file delivery v.s. normalized cache size.

while the solid lines are generated with the default antenna configuration as mentioned before. We can see that the performance with higher antenna number outperforms that with lower number of antennas due to higher antenna array gain. However, the improvement is not significant when compared to gains achieved through WEC, which clearly illustrates the effectiveness of using cache enabled IAB nodes.

Further, while the ASP of file delivery in IBFD mode is superior to that in TDD mode, FDD performs better than IBFD in certain scenarios due to the fact that the FDD mode utilizes orthogonal access and backhaul resources and hence the IAB donor nodes do not produce any interference towards the users. However, for the special case when all files are cached in the IAB nodes, both IBFD and TDD modes outperform FDD as the backhaul transmission is no longer utilized.

However, while throughout the range of cache size, the latency and the throughput in IBFD mode is superior to that in TDD mode, there is a tradeoff between IBFD/TDD mode and FDD with regards to average latency and throughput of file delivery. When the cache size is large, the latency and throughput in IBFD/TDD mode is superior to that in FDD and vice versa for smaller cache size. Moreover, due to the partitioning of bandwidth in FDD mode, caching does not effect the average throughput and latency performance in

FDD based IAB networks. Further, the overall performance of the network irrespective of the transmission mode considering WEC is mostly better than no caching (NC) scenario and MC is superior to UC since MC is more suitable to the skewed content popularity, which is discussed later.

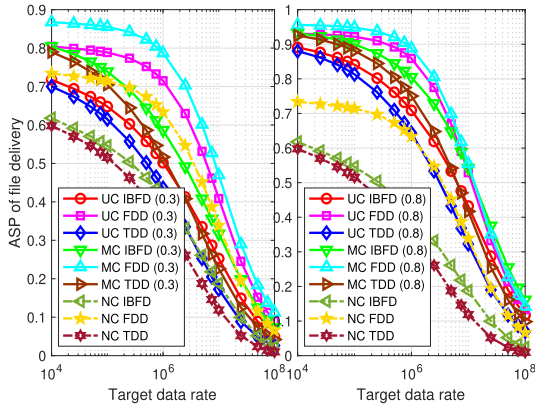


FIGURE 8. ASP of file delivery v.s. target data rate for various normalized cache size.

C. EFFECT OF TARGET DATA RATE

In Fig. 8, we evaluate the ASP of file delivery with respect to the target data rate. When the target rate increases, the probability of successfully delivering the requested file also decreases. Increasing the cache size can improve the ASP of file delivery, which allows mobile operators to provide higher data rates for better quality of experience for users. Further, increasing cache size results in reduction of the performance gap among IBFD, TDD, FDD.

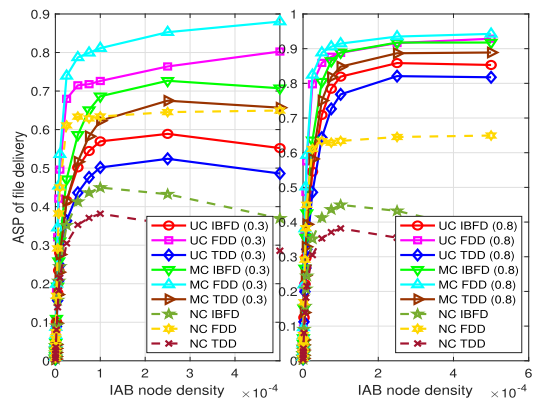


FIGURE 9. ASP of file delivery v.s. IAB node density.

D. EFFECT OF IAB NODE DENSITY

We evaluate the effect of IAB node density on the performance of ASP, latency, and throughput of file delivery in Fig. 9, 10, and 11, respectively. Although higher IAB node density can improve coverage, latency, and throughput performance, the increasing node density results in more interference towards the users, which severely limits the

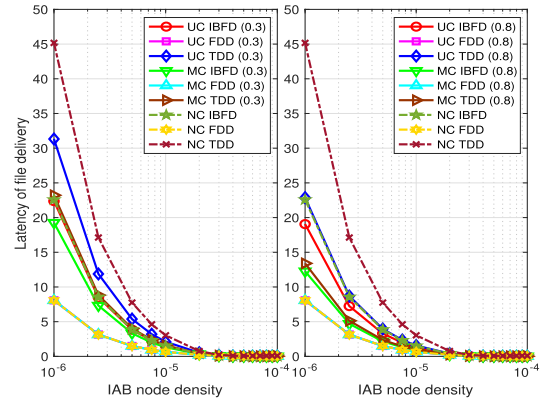


FIGURE 10. Latency of file delivery v.s. IAB node density.

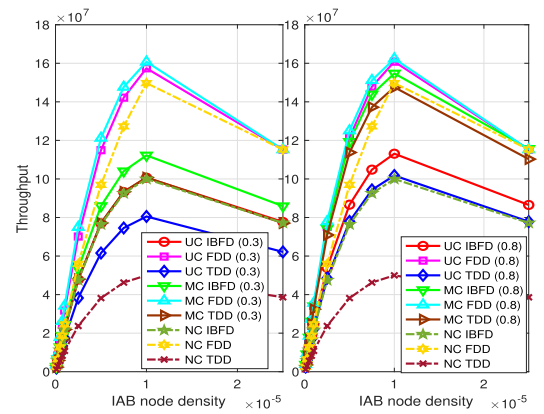


FIGURE 11. Throughput of file delivery v.s. IAB node density.

throughput of file delivery as shown in Fig. 11. However, WEC with large caching capacity is an exciting solution that allows for higher IAB node density for better rate coverage to balance the effect of interference while maintaining (even improving) the latency and throughput performance.

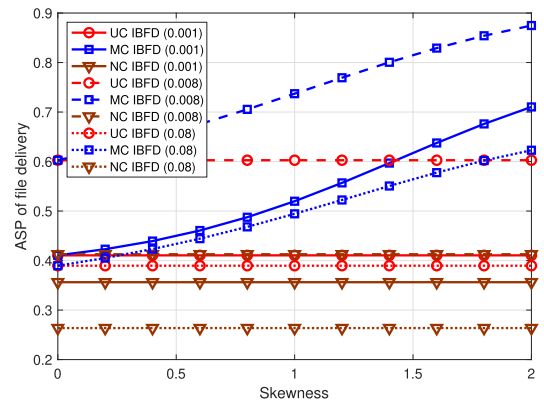


FIGURE 12. ASP of file delivery v.s. skewness with various blockage.

E. EFFECT OF CONTENT POPULARITY

Finally in Fig. 12, 13, and 14, we evaluate the IBFD FR2-IAB network's performance with respect to the caching technique used and the content popularity of files, which is tuned by the

TABLE 2. Computation complexity for calculating some intermediate variables.

Equation	Multiplications	Additions	Total (flops)
$Z(l)$	7	2	9
$\tilde{Z}(l)$	8	3	11
$Z'(l)$	2	–	2
$\tilde{Z}'(l)$	2	1	3
$\lambda_{x^C}(l)$	10	1	11
$\lambda_{x^N}(l)$	9	2	11
\mathcal{A}_s^C [eq. (27)] and \mathcal{A}_m^C [eq. (29)]	34	7	41
\mathcal{A}_s^N [eq. (28)] and \mathcal{A}_m^N [eq. (30)]	33	8	41
eq. (31) and (33)	36	5	41
eq. (32) and (34)	37	8	45
Each conditional ASP for LOS in eq. (35)	118	21	139
Each conditional ASP for NLOS in eq. (35)	119	24	143

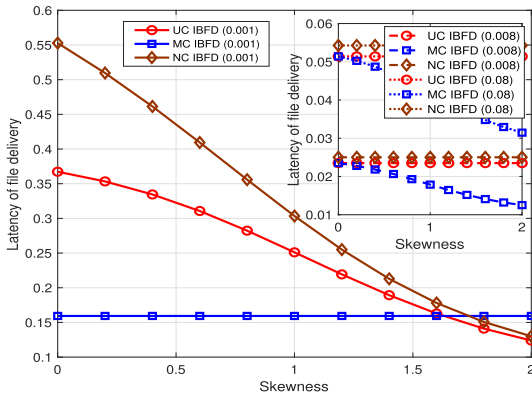


FIGURE 13. Latency of file delivery v.s. skewness with various blockage density.

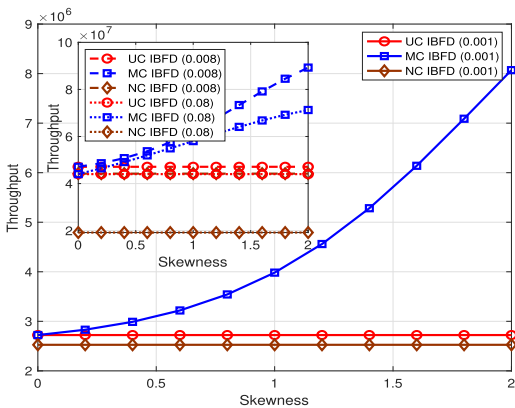


FIGURE 14. Throughput of file delivery v.s. skewness with various blockage density.

skewness parameter. The value inside the bracket signifies the blockage density. When the skewness value is small, the content popularity tends to be uniformly distributed, while a higher value of skewness leads to higher skewed content popularity distribution. This results in the top ranked contents to be highly likely to be requested. Alternatively, when skewness increases, performance of UC remains the same since UC is independent of the content popularity while the performance of MC improves since now it becomes highly possible for most of the popular files that are stored to be requested by all users. Further, we also evaluate the performance of

the network with respect to blockage density. When the blockage density increases, there is a positive effect on the ASP, latency, and throughput of file delivery due to the fact that dense blockages could block certain undesired signals considered as interference at the users. However, if we keep increasing the blockage density, higher propagation loss via NLOS transmission acts as the main limitation that severely degrades the performance of the network.

F. COMPLEXITY OF THE DERIVED PERFORMANCE METRICS

Here, we show the computation complexity of the ASP of file delivery, throughput, and latency in (35), (36), and (41) respectively, by counting the floating-point operations in terms of multiplications and additions (i.e., each multiplication/ addition accounts for 1 flop). However, prior to that we have to first compute the association probabilities given in Proposition 1, the conditional ASP of file delivery that is derived in the Appendix C, and rate calculations in (37)–(40). For simplicity, we assume the integral and exponential operations as one operation (which have linear complexity) and only account for the number of multiplications and additions inside the integral and exponential functions. Accordingly, the approximate complexity analysis of certain intermediate variables in terms of the number of multiplications and additions required is shown in Table 2.

Now, for the case of ASP of file delivery in (35), assuming that the intermediate variables, \mathcal{A}_s^C , \mathcal{A}_s^N , \mathcal{A}_m^C , \mathcal{A}_m^N , and the 4 conditional ASPs of file delivery (i.e., $\mathbb{P}[\mathbf{R}_a^L \geq v]$, etc.) are already calculated (given in Table 2), in each loop we have a total of 13 multiplications and 4 additions. Further, we have a total of F loops, which gives $13F$ multiplications and $5F$ additions, where F is the total number of files. Accordingly, using Table 2, we can give the total number of multiplications and additions required to calculate the ASP of file delivery as $13F + 608F$ (i.e., $118 \times 2 + 119 \times 2 + 34 \times 2 + 33 \times 2$) multiplications and $5F + 120F$ (i.e., $21 \times 2 + 24 \times 2 + 7 \times 2 + 8 \times 2$) additions, respectively. Therefore, a total of ≈ 746 flops are required to compute the ASP of file delivery. Similarly, the computational complexity for throughput can be given as $11F + 612F$ multiplications and $6F + 120F$ additions,

i.e., ≈ 749 flops and for latency as $13F + 612F$ multiplications and $4F + 120F$ additions *i.e.*, ≈ 749 flops.

VI. CONCLUSION

We proposed a 3GPP-inspired analytical framework for a cache-enabled IBFD FR2-IAB network. Utilizing stochastic geometry tools we provided a holistic analysis with respect to three key fundamental network metrics, namely the ASP, latency, and throughput of file delivery. Numerical results show that WEC is an effective solution to not only improve the network performance, but also to alleviate network parameter requirements in terms of the level of RSIC required for IBFD IAB nodes, antenna number, target data rate and IAB node density. Nevertheless, several tradeoffs from the perspective of network design were discovered: 1) WEC reduces the requirement for higher self-interference cancellation. 2) Increasing the cache size at the IAB nodes can improve the ASP of file delivery, which allows mobile operators to provide higher data rates for better quality of experience for users. 3) Latency and throughput performance of cache-enabled IBFD is superior to TDD mode but shows a trade-off when compared to FDD mode, which is highly dependent on cache size. 4) WEC allows to increase the IAB node density by balancing the effect of interference. 5) Increasing the blockage density to a certain extent has a positive effect on the ASP, latency, and throughput of file delivery.

Furthermore, the current analysis does not focus on optimal caching placement schemes. The design of optimal caching schemes is non-trivial and will be considered as an important topic for future research. Similarly, optimal beamforming schemes as shown in [21], [24], [38], that balances between data rate, cache size and latency will be explored in future on top of the current analysis.

**APPENDIX A
PROOF OF LEMMA 1**

We denote the intensity measure of the point process Φ_i with $i \in \{s, m\}$ by its process notation and the thinned process in terms of LOS and NLOS by $\Phi_i^{\mathcal{L}}$ and $\Phi_i^{\mathcal{N}}$ with densities $\lambda_i p_{\mathcal{L}}(r)$ and $\lambda_i p_{\mathcal{N}}(r)$, respectively. Since the superposition of these thinned PPPs is the original PPP, the intensity measure $\Lambda_{\mathcal{L}_i}([0, l])$ with t denoting the location of the associated BS (*i.e.*, $t = x_s$ for $i = s$ and $t = x_m$ for $i = m$) is derived by using the mapping theorem, given as

$$\begin{aligned} \Lambda_{\mathcal{L}_i}([0, l]) &= \Phi_i^{\mathcal{L}}([0, l^{\frac{1}{\alpha_{\mathcal{L}}}}]) + \Phi_i^{\mathcal{N}}([0, l^{\frac{1}{\alpha_{\mathcal{N}}}}]) \\ &= \int_0^{2\pi} \int_0^\infty \lambda_i \exp(-\beta r) \mathbf{1}(r < l^{\frac{1}{\alpha_{\mathcal{L}}}}) r dr d\theta \\ &\quad + \int_0^{2\pi} \int_0^\infty \lambda_i (1 - \exp(-\beta r)) \mathbf{1}(r < l^{\frac{1}{\alpha_{\mathcal{N}}}}) r dr d\theta \\ &= 2\pi \lambda_i \left(Z\left(l^{\frac{1}{\alpha_{\mathcal{L}}}}\right) + \hat{Z}\left(l^{\frac{1}{\alpha_{\mathcal{N}}}}\right) \right), \end{aligned} \tag{A.1}$$

where $Z(\cdot)$ and $\hat{Z}(\cdot)$ are defined in the Lemma 1. Now, the density can be computed by taking derivative of intensity measure with respect to l to complete the proof.

**APPENDIX B
PROOF OF PROPOSITION 1**

Let the distance from the serving IAB node in LOS transmission and NLOS transmission to the typical user be $R_{x_s^{\mathcal{L}0}}$ and $R_{x_s^{\mathcal{N}0}}$, respectively. According to the definition of the association probability, we have

$$\begin{aligned} \mathcal{A}_s^{\mathcal{L}} &= \mathbb{E}_{R_{x_s^{\mathcal{L}0}}} \left[\mathbb{P} \left[R_{x_s^{\mathcal{L}0}}^{-\alpha_{\mathcal{L}}} \geq r^{-\alpha_{\mathcal{N}}} \right] \right] \\ &\stackrel{(a)}{=} \int_0^\infty \mathbb{P} \left[R_{x_s^{\mathcal{L}0}}^{-\alpha_{\mathcal{L}}} \geq r^{-\alpha_{\mathcal{N}}} \right] f_R(R_{x_s^{\mathcal{L}0}}) dR_{x_s^{\mathcal{L}0}} \\ &\stackrel{(b)}{=} \int_0^\infty \mathbb{P} \left[\underbrace{r^{\alpha_{\mathcal{N}}}}_{l_{\mathcal{N}}} \geq \underbrace{R_{x_s^{\mathcal{L}0}}^{\alpha_{\mathcal{L}}}}_{l_{\mathcal{L}}} \right] f_{L_{x_s^{\mathcal{L}}}}(l_{\mathcal{L}}) dl_{\mathcal{L}} \\ &\stackrel{(c)}{=} \int_0^\infty \exp \left[-\Lambda_{\mathcal{L}_{x_s^{\mathcal{N}}}}([0, l_{\mathcal{L}}]) \right] f_{L_{x_s^{\mathcal{L}}}}(l_{\mathcal{L}}) dl_{\mathcal{L}}, \end{aligned} \tag{B.1}$$

where (b) is a revised form of (a) with respect to path loss process and (c) follows from the void probability. In addition, $\Lambda_{\mathcal{L}_{x_s^{\mathcal{N}}}}([0, l_{\mathcal{L}}])$ and $f_{L_{x_s^{\mathcal{L}}}}(l_{\mathcal{L}})$ are given in (24) and (26), respectively. Similarly,

$$\begin{aligned} \mathcal{A}_s^{\mathcal{N}} &= \int_0^\infty \mathbb{P} \left[R_{x_s^{\mathcal{N}0}}^{-\alpha_{\mathcal{N}}} \geq r^{-\alpha_{\mathcal{L}}} \right] f_R(R_{x_s^{\mathcal{N}0}}) dR_{x_s^{\mathcal{N}0}} \\ &= \int_0^\infty \mathbb{P} \left[\underbrace{r^{\alpha_{\mathcal{L}}}}_{l_{\mathcal{L}}} \geq \underbrace{R_{x_s^{\mathcal{N}0}}^{\alpha_{\mathcal{N}}}}_{l_{\mathcal{N}}} \right] f_{L_{x_s^{\mathcal{N}}}}(l_{\mathcal{N}}) dl_{\mathcal{N}} \\ &= \int_0^\infty \exp \left[-\Lambda_{\mathcal{L}_{x_s^{\mathcal{L}}}}([0, l_{\mathcal{N}}]) \right] f_{L_{x_s^{\mathcal{N}}}}(l_{\mathcal{N}}) dl_{\mathcal{N}}. \end{aligned} \tag{B.2}$$

The association probabilities $\mathcal{A}_m^{\mathcal{L}}$ and $\mathcal{A}_m^{\mathcal{N}}$ can be derived in a similar way.

**APPENDIX C
PROOF OF THE ASP OF FILE DELIVERY**

Here, we show how to derive the ASP of file delivery $\mathcal{P}_a^{\mathcal{L}}(v) = \mathbb{P}[R_a^{\mathcal{L}} \geq v]$. Similarly, $\mathbb{P}_a^{\mathcal{N}}(v) = \mathbb{P}[R_a^{\mathcal{N}} \geq v]$, $\mathcal{P}_b^{\mathcal{L}}(v) = \mathbb{P}[R_b^{\mathcal{L}} \geq v]$, and $\mathcal{P}_b^{\mathcal{N}}(v) = \mathbb{P}[R_b^{\mathcal{N}} \geq v]$ can be derived following similar steps.

$$\begin{aligned} \mathcal{P}_a^{\mathcal{L}}(v) &= \mathbb{E} \left\{ \mathbb{P} \left[\text{SINR}_{x_s^{\mathcal{L}}} \geq \underbrace{2^{\frac{v}{W}} - 1}_{\mathcal{Q}} \right] \right\} \\ &\approx \mathbb{E} \left\{ \mathbb{P} \left[\frac{P_s \frac{n_s^m n_r^m}{K U_s \eta_{\mathcal{L}}} |h_{x_s^{\mathcal{L}0}}|^2 |\omega|^2 r_{x_s^{\mathcal{L}0}}^{-\alpha_{\mathcal{L}}} PZF}{I_{s_2}^a + I_m^a + \sigma^2} \geq \mathcal{Q} \right] \right\} \\ &= \int_0^\infty \left(1 - \frac{1}{n_s^m} \right)^{U_s - 1} \mathbb{E} \left\{ \exp \left[-\frac{\mathcal{Q}(I_{s_2}^a + I_m^a + \sigma^2)}{K U_s \frac{n_s^m n_r^m}{\eta_{\mathcal{L}}} \omega^2} \right] \right\} \\ &\quad \times \hat{f}_{L_{x_s^{\mathcal{L}}}}(l_{x_s^{\mathcal{L}0}}) dl_{x_s^{\mathcal{L}0}}, \end{aligned} \tag{C.1}$$

where $l_{x_s^{\mathcal{L}}} = r_{x_s^{\mathcal{L}}}^{\alpha_{\mathcal{L}}}$. In particular, the expectation above is given as

$$\mathbb{E} \left\{ \exp \left[- \frac{Q\sigma^2 U_s K l_{x_s^{\mathcal{L}}} \eta_{\mathcal{L}}}{\omega^2 P_s n_t^s n_r^u} \right] \right\} \underbrace{\mathbb{E} \left\{ \exp \left[- \frac{Q I_{s_2}^a U_s K l_{x_s^{\mathcal{L}}} \eta_{\mathcal{L}}}{\omega^2 P_s n_t^s n_r^u} \right] \right\}}_A \times \underbrace{\mathbb{E} \left\{ \exp \left[- \frac{Q I_m^a U_s K l_{x_s^{\mathcal{L}}} \eta_{\mathcal{L}}}{\omega^2 P_s n_t^s n_r^u} \right] \right\}}_B. \quad (C.2)$$

In the following, we calculate the expectation terms, A and B , respectively. In particular, due to the thinning theorem with respect to LOS and NLOS,

$$A = \mathbb{E} \left\{ \exp \left[- \frac{Q U_s K l_{x_s^{\mathcal{L}}} \eta_{\mathcal{L}}}{\omega^2 n_t^s n_r^u P_s} (I_{s_2}^{\mathcal{L}} + I_{s_2}^{\mathcal{N}}) \right] \right\} \\ = \underbrace{\mathbb{E} \left\{ \exp \left[- \frac{Q U_s K l_{x_s^{\mathcal{L}}} \eta_{\mathcal{L}}}{\omega^2 n_t^s n_r^u P_s} I_{s_2}^{\mathcal{L}} \right] \right\}}_{A_1} \\ \times \underbrace{\mathbb{E} \left\{ \exp \left[- \frac{Q U_s K l_{x_s^{\mathcal{L}}} \eta_{\mathcal{L}}}{\omega^2 n_t^s n_r^u P_s} I_{s_2}^{\mathcal{N}} \right] \right\}}_{A_2}. \quad (C.3)$$

Now, let $s_{\mathcal{L}} = \frac{Q U_s K l_{x_s^{\mathcal{L}}} \eta_{\mathcal{L}}}{\omega^2 n_t^s n_r^u P_s}$ and substitute it into the above function. Then,

$$A_1 = \mathbb{E} \left\{ \exp \left[- s_{\mathcal{L}} \sum_{\substack{b \in \Phi_s^{\mathcal{L}}, \\ b \neq x_s^{\mathcal{L}}}} \frac{P_s n_t^s n_r^u}{l_{b0} \eta_{\mathcal{L}} \bar{U}_s K} \sum_{u \in \mathcal{U}_b} \left| \sum_{i=1}^{\eta_{\mathcal{L}}} h_{i b0} \omega \gamma_{i a s} \right|^2 \right] \right\} \quad (C.4)$$

$$\stackrel{(a)}{\geq} \mathbb{E} \left\{ \prod_{\substack{b \in \Phi_s^{\mathcal{L}}, \\ b \neq x_s^{\mathcal{L}}}} \mathbb{E} \left\{ \exp \left[- s_{\mathcal{L}} \frac{P_s n_t^s n_r^u}{l_{b0} \eta_{\mathcal{L}} \bar{U}_s K} \left(\sum_{u \in \mathcal{U}_b} \sum_{i=1}^{\eta_{\mathcal{L}}} (\gamma_{i a s})^2 \right) \right] \right\} \right. \\ \left. \times \sum_{i=1}^{\eta_{\mathcal{L}}} |h_{i b0}|^2 \omega^2 \right\} \quad (C.5)$$

$$\stackrel{(b)}{=} \mathbb{E} \left\{ \prod_{\substack{b \in \Phi_s^{\mathcal{L}}, \\ b \neq x_s^{\mathcal{L}}}} \left(1 + s_{\mathcal{L}} \frac{P_s n_t^s n_r^u}{l_{b0} \eta_{\mathcal{L}} \bar{U}_s K} \left(\sum_{u \in \mathcal{U}_b} \sum_{i=1}^{\eta_{\mathcal{L}}} (\gamma_{i a s})^2 \right) \omega^2 \right)^{-\eta_{\mathcal{L}}} \right\} \quad (C.6)$$

$$\stackrel{(c)}{\geq} \exp \left\{ - \int_{l_{x_s^{\mathcal{L}}}}^{\infty} \left[1 - \left(1 + \frac{\omega^2}{K} s_{\mathcal{L}} P_s n_t^s n_r^u l_{b0}^{-1} \right)^{-\eta_{\mathcal{L}}} \right] \right. \\ \left. \times \lambda_{\mathcal{L}, x_s^{\mathcal{L}}} (l_{b0}) dl_{b0} \right\}, \quad (C.7)$$

where (a) follows from that Cauchy-Schwarz inequality. (b) follows from the fact that $\sum_{i=1}^{\eta_{\mathcal{L}}} |h_{i b0}|^2$ follows Chi-square/gamma distribution with parameters $\eta_{\mathcal{L}}$ and 1. (c) follows from the fact that $\gamma_{i a s}^2 \leq 1$ and the probability generating functional of a PPP. In a similar way, the other expectation term A_2 is lower bounded. Similarly, we can compute the expectation B but the density of the interfering IAB donor

nodes should be scaled by a factor $p_{\text{miss}} = 1 - p_{\text{hit}}^{U_s}$, which means that IAB donor nodes will serve IAB nodes if there are cache miss users. In a similar way, the other conditional ASPs of file delivery in (35) can be computed.

REFERENCES

- [1] J. Liu, M. Sheng, L. Liu, and J. Li, "Network densification in 5G: From the short-range communications perspective," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 96–102, Dec. 2017.
- [2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
- [3] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.
- [4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [5] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2196–2211, Oct. 2015.
- [6] J. Ghimire and C. Rosenberg, "Revisiting scheduling in heterogeneous networks when the backhaul is limited," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2039–2051, Oct. 2015.
- [7] *NR: Study on Integrated Access and Backhaul*, 3GPP, document TR 38.874, 2017.
- [8] C. Saha, M. Afshang, and H. S. Dhillon, "Bandwidth partitioning and downlink analysis in millimeter wave integrated access and backhaul for 5G," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8195–8210, Dec. 2018.
- [9] M. Polese, M. Giordani, T. Zugno, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated access and backhaul in 5G mmWave networks: Potentials and challenges," 2019, *arXiv:1906.01099*. [Online]. Available: <http://arxiv.org/abs/1906.01099>
- [10] W. Pu, X. Li, J. Yuan, and X. Yang, "Resource allocation for millimeter wave self-backhaul network using Markov approximation," *IEEE Access*, vol. 7, pp. 61283–61295, 2019.
- [11] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.
- [12] G. S. Paschos, G. Iosifidis, M. Tao, D. Towsley, and G. Caire, "The role of caching in future communication systems and networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1111–1125, Jun. 2018.
- [13] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.
- [14] E. Bastug, M. Bennis, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2014, pp. 649–653.
- [15] D. Liu and C. Yang, "Caching policy toward maximal success probability and area spectral efficiency of cache-enabled HetNets," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2699–2714, Jun. 2017.
- [16] N. Giatsoyglou, K. Ntontin, E. Kartsakli, A. Antonopoulos, and C. Verikoukis, "D2D-aware device caching in mmWave-cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2025–2037, Sep. 2017.
- [17] L. Wang, K.-K. Wong, S. Lambotharan, A. Nallanathan, and M. ElKashlan, "Edge caching in dense heterogeneous cellular networks with massive MIMO-aided self-backhaul," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6360–6372, Sep. 2018.
- [18] S. Biswas, T. Zhang, K. Singh, S. Vuppala, and T. Ratnarajah, "An analysis on caching placement for millimeter-micro-wave hybrid networks," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1645–1662, Feb. 2019.
- [19] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

- [20] A. Cirik, "Beamforming design for full-duplex MIMO interference channels—QoS and energy-efficiency considerations," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4635–4651, Nov. 2016.
- [21] Z. Chu, F. Zhou, P. Xiao, Z. Zhu, D. Mi, N. Al-Dhahir, and R. Tafazolli, "Resource allocation for secure wireless powered integrated multicast and unicast services with full duplex self-energy recycling," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 620–636, Jan. 2019.
- [22] X. Xia, K. Xu, Y. Wang, and Y. Xu, "A 5G-enabling technology: Benefits, feasibility, and limitations of in-band full-duplex mMIMO," *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 81–90, Sep. 2018.
- [23] X. Xia, K. Xu, D. Zhang, Y. Xu, and Y. Wang, "Beam-domain full-duplex massive MIMO: Realizing co-time co-frequency uplink and downlink transmission in the cellular system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8845–8862, Oct. 2017.
- [24] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted sum rate maximization in full-duplex multi-user multi-cell MIMO networks," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1590–1608, Apr. 2017.
- [25] C. Saha and H. S. Dhillon, "Millimeter wave integrated access and backhaul in 5G: Performance analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 12, pp. 2669–2684, Dec. 2019.
- [26] J. Lee and T. Q. S. Quek, "Hybrid Full-/Half-Duplex system analysis in heterogeneous wireless networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2883–2895, May 2015.
- [27] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "A comparison of MIMO techniques in downlink millimeter wave cellular networks with hybrid beamforming," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1952–1967, May 2016.
- [28] S. H. Chae and W. Choi, "Caching placement in stochastic wireless caching helper networks: Channel selection diversity via caching," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6626–6637, Oct. 2016.
- [29] Z. Chen, J. Lee, T. Q. S. Quek, and M. Kountouris, "Cooperative caching and transmission design in cluster-centric small cell networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3401–3415, May 2017.
- [30] A. Thornburg, T. Bai, and R. W. Heath, "Performance analysis of outdoor mmWave ad hoc networks," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 4065–4079, Aug. 2016.
- [31] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.
- [32] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association with traditional macrocells and millimeter wave small cells," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6244–6258, Sep. 2016.
- [33] M. N. Kulkarni, A. Alkhateeb, and J. G. Andrews, "A tractable model for per user rate in multiuser millimeter wave cellular networks," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2015, pp. 328–332.
- [34] A. Sharma, R. K. Ganti, and J. K. Milleth, "Joint backhaul-access analysis of full duplex self-backhauling heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1727–1740, Mar. 2017.
- [35] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.
- [36] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 167–203, 1st Quart., 2017.
- [37] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [38] Z. Zhu, S. Huang, Z. Chu, F. Zhou, D. Zhang, and I. Lee, "Robust designs of beamforming and power splitting for distributed antenna systems with wireless energy harvesting," *IEEE Syst. J.*, vol. 13, no. 1, pp. 30–41, Mar. 2019.



TONG ZHANG received the M.Sc. degree in signal processing and communications from The University of Edinburgh, Edinburgh, U.K., in 2016, where he is currently pursuing the Ph.D. degree with the Institute for Digital Communications. His research interests include wireless edge caching, millimeter-wave communications, and stochastic geometry.



SUDIP BISWAS (Member, IEEE) received the Ph.D. degree in digital communications from The University of Edinburgh (UEDIN), U.K., in 2017. He held the position of Research Associate at the Institute for Digital Communications, UEDIN, from 2017 to 2019. He also has industrial experience at Tata Consultancy Services, India (Lucknow and Kolkata), where he held the position of an Assistant Systems Engineer, from 2010 to 2012. He currently works with the Department of Electronics and Communications Engineering, Indian Institute of Information Technology Guwahati (IIITG), as an Assistant Professor. He leads research on signal processing for wireless communications, with a particular focus on 5G's long-term evolution, including transceiver design for full-duplex radios, wireless edge caching, communications-radar co-existence, and intelligent reflector surface assisted communication. He was an Organizer of the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Edinburgh, U.K., in 2016, and has been involved in EU FP7 projects such as the Remote Radio Heads and Parasitic Antenna Arrays (HARP) and the Dynamic Licensed Shared Access (ADEL), a DST UKIERI project on wireless edge caching and an EPSRC project on NoMA.



THARMALINGAM RATNARAJAH (Senior Member, IEEE) was the Head of the Institute for Digital Communications, from 2016 to 2018. Prior to this, he held various positions at McMaster University, Hamilton, ON, Canada, from 1997 to 1998; Nortel Networks, Ottawa, Canada, from 1998 to 2002; the University of Ottawa, Canada, from 2002 to 2004; and Queen's University Belfast, U.K., from 2004 to 2012. He is currently with the Institute for Digital Communications,

The University of Edinburgh, Edinburgh, U.K., as a Professor in digital communications and signal processing. He has supervised 15 Ph.D. students and 20 postdoctoral research fellows, and raised more than 11 million USD of research funding. He was the Coordinator of the EU projects ADEL (3.7M€) in the area of licensed shared access for 5G wireless networks, HARP (4.6M€) in the area of highly distributed MIMO, the EU Future and Emerging Technologies projects HIATUS (3.6M€) in the area of interference alignment, and CROWN (3.4M€) in the area of cognitive radio networks. His research interests include signal processing and information theoretic aspects of 5G and beyond wireless networks, full-duplex radio, mmWave communications, random matrix theory, interference alignment, statistical and array signal processing, and quantum information theory. He has published over 400 articles in these areas and holds four U.S. patents. He is a member of the American Mathematical Society and the Information Theory Society and a Fellow of the Higher Education Academy (FHEA). He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, from 2015 to 2017, and the Technical Co-Chair in the 17th IEEE International Workshop on Signal Processing Advances in Wireless Communications, Edinburgh, in 2016.

...