

# An Analysis on Caching Placement for Millimetre-Micro Wave Hybrid Networks



# An Analysis on Caching Placement for Millimetre-Micro Wave Hybrid Networks

Sudip Biswas, Tong Zhang, Keshav Singh,

Satyanarayana Vuppala, and Tharmalingam Ratnarajah

**Abstract**

In this paper, we investigate the feasibility of *wireless edge caching* in a hybrid millimeter wave (mmWave)–micro wave ($\mu$Wave) network. Considering the average success probability (ASP) of file delivery as the performance metric, we derive expressions for the association probability of the typical user to the mmWave and $\mu$Wave networks using stochastic geometric tools. Accordingly, we provide an upper bound on the ASP of file delivery and formulate the content caching placement scheme as an optimization problem with respect to caching probabilities, that jointly optimizes the ASP of file delivery considering both content placement and delivery phases. To simplify the non-convex problem and obtain design insights, we split it into two scenarios: i) noise-limited and ii) interference-limited, and then propose optimal caching placement algorithms for both. We numerically evaluate the performance of the proposed schemes under several essential factors, such as content popularity, cache size, target data rate, blockages in the mmWave network, BS density, and path loss and also compare it with other common proactive caching schemes, namely uniform caching, caching $M$ most popular files, and random caching. Numerical results demonstrate the superiority of the proposed caching scheme over others, albeit certain trade-offs.

**Index Terms**

Wireless edge caching, hybrid millimetre-micro wave network, Poison point processes.

## I. Introduction

Wireless network operators around the world are faced with the conundrum of complying with the growing demands of quality of service (QoS) requirements of users. The primary reason for this is the proliferation of wireless devices and services, which has led to an exaggeration of mobile data traffic to the point that network providers are seeking for alternative solutions to the

Sudip Biswas, Tong Zhang, Keshav Singh, and Tharmalingam Ratnarajah are with Institute for Digital Communications, School of Engineering, University of Edinburgh, Edinburgh, UK. Email: {sudip.biswas, k.singh, t.zhang, t.ratnarajah}@ed.ac.uk.

Satyanarayana Vuppala is with University of Luxembourg, Luxembourg. Email: satyanarayana.vuppala@uni.lu.

currently available service schemes. Based on the well known network throughput measurement formula as given below,

$$\underbrace{\text{Throughput}}_{\text{bits/s/km}^2} = \underbrace{\text{Cell density}}_{\text{cell/km}^2} \times \underbrace{\text{Available spectrum}}_{\text{Hz}} \times \underbrace{\text{Spectrum efficiency}}_{\text{bits/s/Hz/cell}},$$

two potential solutions towards improving wireless network throughput can be identified as: 1) network densification through small cells and heterogenous networks, and 2) moving to mmWave frequency bands. However, while above solutions are beneficial for the access links, they do little to alleviate the burden on the backhaul links, which is further exaggerated when heterogeneous networks are considered. In a typical cellular network, base stations (BSs) retrieve requested files using capacity-limited backhaul links. During peak hours, this results in an information-congestion bottleneck both at the BSs as well as in the backhaul links. Nonetheless, it is interesting to note that a substantial amount of the data traffic are redundantly generated over networks [1], as several popular contents are asynchronously and repeatedly requested by many users. Motivated by this, pre-fetching some popular contents in the local caches of base stations (BSs), also termed as wireless edge caching, can alleviate network backhaul traffic loads, by which the requested content will be served directly to the users by one of the neighbouring BSs depending on the availability of the file in its local cache and the association criteria of the users to the BSs. Additionally, wireless edge caching also has the advantages of $1)$ reducing latency by shortening the communication distance, $2)$ improving network capacity and throughput, and $3)$ reducing operational cost due to lower cost of storage memory than bandwidth.

The above discussion clearly adds up to the fact that the solutions provided for access along with wireless edge caching strive towards fulfilling common goals of improving the QoS of networks and quality of experience (QoE) for users, which makes it imperative to investigate the performance of these technologies in a co-existed framework. While recent studies focus on the performance analysis of either mmWave networks [2]–[8] or hybrid networks [9]–[15], or wireless caching [16]–[25], to the best of authors' knowledge, at the time of writing this paper, no work considers the amalgamation of all. Consequently, in this work, we consider a hybrid wireless network architecture consisting of multiple small and macro cells, equipped respectively with mmWave and micro wave ($\mu$Wave) BSs, that consist of storage memories to cache popular contents. While the $\mu$Wave network can be looked upon as a conventional sub-6 GHz UHF based 3G and 4G network, the mmWave spectrum spans between $30 - 300$ GHz. Further, the

consideration of mmWave transmission for small cells and $\mu$Wave transmission for macro cells is justified by virtue of the fact that unlike $\mu$Wave signals, which can propagate through long distances, mmWave signals due to its short wavelengths have limited transmission range and are easily attenuated by blockages, such as buildings, trees, etc. Accordingly, synchronicity between the both could play a vital role in providing wider coverage, particularly in sparse deployment of mmWave networks.

Motivated by the above discussion, in this paper, we investigate the optimal content caching placement strategy in a mm/$\mu$Wave hybrid network by maximizing the average success probability (ASP) of file delivery. Note that the consideration of optimal caching placement in a hybrid wireless network introduces fundamental new challenges, the solution of which requires rigorous optimization and analysis and is addressed in this paper. The main distinctions are summarized in terms of design guidelines as follows:

1) Considering a stochastic geometric framework, we model a hybrid two-tier cellular network involving two radio access technologies, namely mmWave and $\mu$Wave.
2) We calculate the association probability of the user to a mmWave or $\mu$Wave network based on the long-term average biased (LTAB) received signal power, details of which are provided in later sections of the paper.
3) Using the association probability, we provide a holistic analytical expression for the ASP of file delivery. While, a closed-form expression is provided for the noise-limited scenario, an upper bound is provided for the interference-limited case.
4) Next, to place the contents in the hybrid network, we maximize the ASP of file delivery by considering finite memory size and content popularity. This is achieved by optimally computing the caching probabilities in the file caching placement phase. Note that the computation for the interference limited scenario involves solving of a non-trivial difference of convex (DC) problem.
5) To implement the above efficiently, we propose two algorithms, one each for noise-limited and interference-limited scenarios. Both algorithms are tested numerically and compared with existing caching placement schemes such as, $1)$ uniform caching placement, $2)$ caching $M$ most popular contents, and $3)$ random caching placement.
6) Finally, to obtain design insights, we evaluate the effects of several essential parameters,

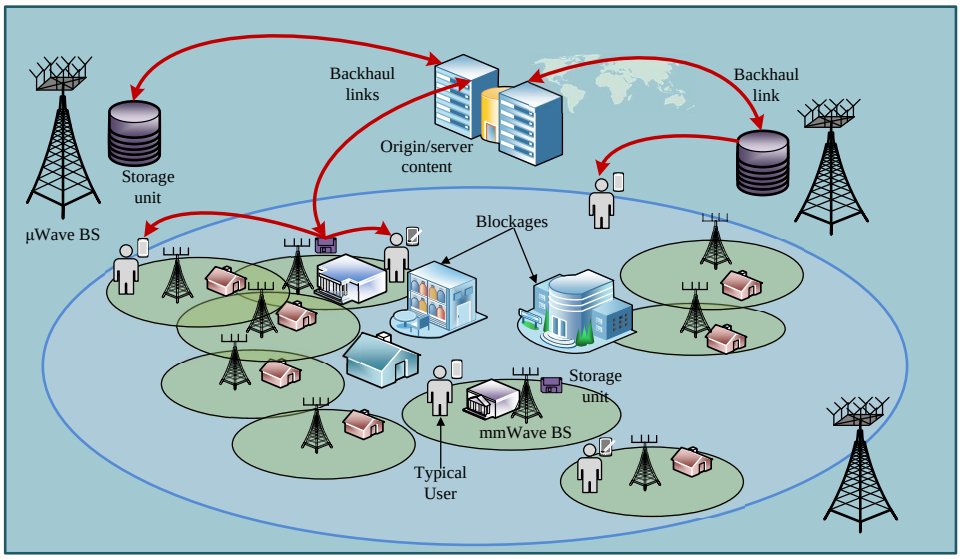


Fig. 1: An illustration of a cache enabled hybrid mmWave-$\mu$Wave network.

such as content popularity, cache size, target data rate, mm/$\mu$Wave BS density, blockages in the network, and path loss on the ASP of file delivery.

## II. System Model

We consider the downlink transmission in a cache-enabled two-tier hybrid cellular network comprising of small cell network (SCN) and macro cell network (MCN) as shown in Fig. 1. While mmWave BSs and associated users form the SCN, $\mu$Wave BSs and their associated users form the MCN. The mmWave BSs and $\mu$Wave BSs are independently modeled by two independent homogeneous Poisson point processes (PPPs) $\Phi_m$ with density $\lambda_m$ and $\Phi_\mu$ with density $\lambda_\mu$, respectively. The users in the network follow another independent homogenous PPP $\Phi_u$ with density $\lambda_u$. Since the set of BSs or users belonging to a particular network (SCN or MCN) operate in the same spectrum (mm or $\mu$Wave), it does not interfere with the set of BSs or users of the other network. Further, all the mmWave and $\mu$Wave BSs are equipped with multiple antennas $n_t^m$ and $n_t^\mu$, respectively. The users are assumed to be equipped with two sets of RF chains with antennas $n_r^m$ and $n_r^\mu$ to receive mmWave and $\mu$Wave signals, respectively.

*Remark 1:We assume $n_r^m > 1$ and $n_r^\mu = 1$. This is due to the intrinsic relation between wavelength of signals and antenna separation, whereby the wavelength of $\mu$Wave signals is much larger than mmWave signals and hence, much larger separation is required between antennas for $\mu$Wave to avoid correlation and coupling. As a result, accommodating more than one antenna at small devices, like mobile phones operating in the $\mu$Wave spectrum may not be feasible.*

The analysis hereinafter is performed for the typical user[1] located at the origin, who will be associated to either the mmWave or $\mu$Wave network depending on a specific association policy. For simplicity, this work considers the best-case scenario where the file of interest requested by the typical user can always be found in the local caches of its associated BS, which enables the optimal caching placement to achieve an ideal state. In the following, we elaborate on the considered system model. Unless otherwise stated, throughout the paper subscript/superscript/notation of $m$ and $\mu$ will be used to refer to mmWave and $\mu$Wave entities, respectively.

### *A. Caching model*

We consider a proactive caching model, wherein to cache popular files requested by a user, each BS contains storage units, referred to as local caches. Additionally, a central source/server containing a global cache[2] is accessible to all the BSs in the hybrid network via wired backhaul links. For simplicity, we represent the size of the cache by the number of files. It is assumed that each mmWave and $\mu$Wave BS can cache up to $\mathrm{C}_m$ and $\mathrm{C}_\mu$ files of length $S$ bits each[3], respectively, such that $\mathrm{C}_m < \mathrm{C}_\mu$. Further, we assume that the distribution of users' requests follows the independent reference model, in which the content popularity is stationary and each user independently requests a data-file from the file set $\mathcal{S} \triangleq \{s_1, s_2, \ldots, s_i, \ldots s_L\}$, where $L$ is the total number of files cached in the network. The popularity of the requested files is assumed to be independent of each other and is modelled by the Zipf distribution [17]. In particular, the popularity of the $i$th file in the library is given as $f_i = (1/i^{\upsilon})/\sum_{j=1}^{L}(1/j^{\upsilon}),\ 1 \leq i \leq L$, where $\upsilon$ is the Zipf exponent, that controls the skewness of the content popularity. The Zipf distribution is considered in line with recent literature [17], [26], where it was shown to be the most suitable distribution for video files. However, other models can also be used and can be expected to exhibit similar trends for the proposed optimal schemes.

Now, to place the contents, a probabilistic caching strategy is assumed, where each BS (mm or $\mu$Wave) caches its file in an independently and identically distributed (i.i.d) manner by producing $C_j$ indices generated according to the distribution $\Pi_j \triangleq \{p_{j_i} : s_i \in \mathcal{S}, i = 1, 2, ..., L\}$, where $p_{j_i}$ is the probability of caching the $i$th file $s_i \in \mathcal{S}$ such that $0 \leq p_{j_i} \leq 1$ and $\sum_{i=1}^{L} p_{j_i} \leq C_j$ with

[1]Slivnyak's theorem enables the characterization of the network performance through the performance of the typical user.

[2]In the event that the file requested by the typical user is not in the local cache, then the file is retrieved from global cache that contains all the files a user may request.

[3]For analytical simplicity, we assume that the size of each file is identical and normalized to one. In the case 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.

$j \in \{m, \mu\}$. The files are proactively cached in advance during off-peak hours through prior requests or overhearing. Now, according to the thinning theorem, the BSs storing the file $i$ are further modelled as an independent PPP with density $\lambda_{j_i} \triangleq p_{j_i}\lambda_j$. Now, from the available $L$ files, the typical user requests one file depending on the file popularity $f_i$. A file with higher popularity is assumed to be requested with higher likelihood. For analytical tractability, hereinafter we assume that the popularity of the files is perfectly known and stationary. This assumption is perhaps over simplistic, but we leave the investigation of unknown and time-varying popularity to future work.

### *B. Hybrid network model [SCN]*

*1) Blockage model:* MmWave signals in the SCN are susceptible to blockages, making it imperative to model blockages for true representation of such systems. We consider the blockages to be stationary blocks which are invariant with respect to direction. We adopt the modeling of blockages in [27], and accordingly, consider a two state statistical model for each link. The link can be either line-of-sight (LOS) [$\mathcal{L}$] or non-line-of-sight (NLOS) [$\mathcal{N}$]. Let the LOS link be of length $r$ and $\beta$ be the blockage density, then the probabilities of occurrence $p_{\mathcal{L}}(.)$ and $p_{\mathcal{N}}(.)$ of LOS and NLOS states, respectively, can be given as a function of $r$ as $p_{\mathcal{L}}(r) = e^{-\beta r}$, $p_{\mathcal{N}}(r) = 1 - e^{-\beta r}$.

*2) Beamforming model:* Due to the small wavelengths of mmWave signals, directional beamforming at both transmitters and receivers can be exploited for compensating path loss and additional noise. The beam patterns are approximated as sectorized gain patterns [28]. Accordingly, the antenna gain pattern for a transmit or receive node about an angle $\phi$ is given as

$$G_q(\theta) = \left\{ \begin{array}{ll} G & \text{if} |\phi| \leq \theta \\ g & \text{if} |\phi| > \theta \end{array} \right\}, \tag{1}$$

where $\theta$ is the beamwidth of the main lobe, $q \in \{\mathrm{T}, \mathrm{R}\}$, with $\mathrm{T}$ denoting the transmitter, and $\mathrm{R}$ the receiver, $\phi \in [0, 2\pi)$ is the angle of boresight direction and $G$ and $g$ are the array gains of main and side lobes, respectively. Hence, the effective antenna gain/interference $G_x$ seen by a user from a BS $x \in \Phi_m$ depends on the directivity of the gains of main (*i.e.,* $G$) and side (*i.e.,* $g$) lobes of the antenna beam pattern, given as

$$G_x = \left\{ \begin{array}{lll} GG, & \text{w.p.} & (\frac{\theta}{2\pi})^2 \\ Gg, & \text{w.p.} & \frac{2\theta(2\pi-\theta)}{(2\pi)^2} \\ gg, & \text{w.p.} & (\frac{2\pi-\theta}{2\pi})^2 \end{array} \right\}. \tag{2}$$

*3) Channel model:* To capture a generalized propagation environment and for analytical tractability, in this work we consider the Nakagami fading model, with $\hat{m}$ being the Nakagami fading parameter and $\Gamma(\hat{m})$ the gamma function. This choice is motivated by the use of this model to simulate small scale fading in recent literature [27], [28].

### C. Hybrid network model [MCN]

The $\mu$Wave channels in the MCN are modeled in a similar way as that of their mmWave counterparts with the only exceptions that the antennas are now omnidirectional with transmit signal power $\mathrm{P}_\mu$ and path loss exponent $\alpha_\mu$. Note that the blockage effects are not considered for $\mu$Wave systems due to low penetration loss of $\mu$Wave signals. Under the consideration of separate encoding scheme at each BS, the BS $x$ sends an information symbol $s_x$ through a linear beamforming vector $\mathbf{v}_x = [v_x^1, \ldots, {v_x}^{n_t^\mu}]^T$, $x \in \Phi_\mu$. Now, by a slight abuse of notation, the received signal at the typical user served by the $\mu$Wave BS $x$ can be given as

$$y_0 = \sqrt{\mathrm{P}_\mu}\mathbf{h}_{0,x}^H\mathbf{v}_{0,x}r_{0,x}^{-\alpha_\mu/2}s_{0,x} + \sqrt{\mathrm{P}_\mu}\sum_{k\in\mathcal{U}_x\setminus\{0\}}\mathbf{h}_{0,x}^H\mathbf{v}_{k,x}r_{0,x}^{-\alpha_\mu/2}s_{k,x} + \sqrt{\mathrm{P}_\mu}\sum_{w\in\Phi_\mu\setminus\{x\}}\sum_{k'\in\mathcal{U}_w}\mathbf{h}_{0,w}^H\mathbf{v}_{k',w}r_{0,w}^{-\alpha_\mu/2}s_{k',w} + n_{0,x}, \tag{3}$$

where $\mathbb{E}[|s_x|^2] = 1$, $\mathbf{h}_{0,x} = [h_{0,x}^1, \ldots, h_{0,x}^{n_t^\mu}]^T \subseteq \mathbf{H}_x$ is the channel between the $\mu$Wave BS $x$ and the typical user and each entry of $\mathbf{h}_{0,x}$ is i.i.d. according to complex normal distribution. Further, $\mathbf{H}_x$ is the channel matrix formed by the channels of all the users associated to the BS $x$, which is denoted by the set $\mathcal{U}_x$ and $n_{0,x}$ denotes the additive Gaussian noise seen at the typical user with zero mean and variance $\sigma_\mu^2$.

## III. Rate Characterization for the Typical User

The typical user may be associated to either a mmWave BS or a $\mu$Wave BS depending on an association policy, which will be discussed in the next section. In this section, we characterize the rate of the typical user when connected to either a mmWave or a $\mu$Wave BS. Let $\Phi_{m_i}$ be the set of mmWave BSs, which have the file $i$ in their local caches. Given that a typical user is associated to a mmWave BS $x_i$ (from $\Phi_{m_i}$) that contains the requested file $i$, the strongest received signal power is given as

$$\zeta_{m_{0,x_i}} = \max_{x_i\in\Phi_{m_i}}\left\{\frac{\mathrm{P}_m G_{0,x_i}\mathcal{X}_{0,x_i}}{r_{0,x_i}^{\alpha_j}}\right\}, \tag{4}$$

where $\mathrm{P}_m$, $G_{0,x_i}$ and $\mathcal{X}_{0,x_i}$ are the transmit power, effective directional antenna gain, and channel power coefficient at the typical user, respectively. In the above, $r_{0,x_i}$ is the distance between the typical user and the serving mmWave BS. Now, let $\Phi_{m_i}^c$ (or $\overline{\Phi}_{m_i}$) be the set of BSs that do not store the file $i$ in their cache memory. Then, $\Phi_{m_i}^c$ is written as $\Phi_{m_i}^c = \Phi_m\setminus\Phi_{m_i}$. Further, for simplicity, perfect beam alignment between the mmWave BS $x_i$ and its associated users is

considered[4] [29]. Accordingly, interference from signals transmitted to other associated users ($\mathcal{U}_{x_i}\backslash\{0\}$) with $x_i \in \Phi_{m_i}$ can be neglected due to the directivity of the beams. Hence, the SINR at the typical user, receiving the file $i$ from the mmWave BS $x_i$ can be defined as

$$\gamma_{m_{0,x_i}} \triangleq \frac{\mathrm{P}_m G_{0,x_i} \mathcal{X}_{0,x_i} r_{0,x_i}^{-\alpha_j}}{\left\{ \underbrace{\sigma_m^2}_{A} + \underbrace{\sum\nolimits_{t\in\Phi_m\backslash\{x_i\}} \mathrm{P}_m G_{0,t} \mathcal{X}_{0,t} r_{0,t}^{-\alpha_j}}_{B} \right\}}. \tag{5}$$

In the denominator above, $A$ represents the noise power at the typical user, and $B$ is the interference seen by the typical user from all other mmWave BSs except for the associated BS. Hereinafter, for notational simplicity, we omit the subscript $0$ used to represent the typical user. Accordingly, the average downlink data rate at the typical user can be given as

$$\mathcal{R}_{m_{x_i}} = (1/N_{x_i}) \times \log_2\left(1+\gamma_{m_{x_i}}\right), \tag{6}$$

where $N_{x_i}$ is the cell load at the serving BS.

Similarly, for the scenario that a user is associated with a $\mu$Wave BS, the downlink rate at the typical user, receiving file $i$ is given as

$$\mathcal{R}_{\mu_{x_i}} = \frac{1}{N_{x_i}} \times \log_2\left(1+\gamma_{\mu_{x_i}}\right), \text{ where } \gamma_{\mu_{x_i}} \triangleq \frac{\mathrm{P}_\mu H_{x_i} r_{x_i}^{-\alpha_\mu}}{\left\{ \underbrace{\sigma_\mu^2}_{A} + \underbrace{\sum\nolimits_{t\in\Phi_\mu\backslash x_i} \mathrm{P}_\mu H_t r_t^{-\alpha_\mu}}_{B} \right\}}. \tag{7}$$

The above rate is obtained by applying a zero-forcing (ZF) precoding scheme at the serving $\mu$Wave BS $x_i$, with $x_i \in \Phi_{\mu_i}$, wherein interference from signals transmitted to other users in (3) is assumed to be cancelled. In the above, $H_{x_i} = ||\mathbf{h}_{x_i}^H \mathbf{v}_{x_i}||^2$ denotes the effective power gain[5] , with $H_{x_i} \sim \Gamma(n_t^\mu - N_{x_i} + 1, 1)$ and $\mathbf{v}_{x_i} \subseteq \mathbf{V}_{x_i} = \mathbf{H}_{x_i}(\mathbf{H}_{x_i}^H \mathbf{H}_{x_i})^{-1}$ is the ZF precoding vector meant for the typical user. Like before, $A$ and $B$ denote noise and interference from all other $\mu$Wave BSs except for the associated $\mu$Wave BS, respectively.

## IV. Association Policy

In this section, we present the communication policy, where we determine the association probabilities $p_{mw}$ and $p_{\mu w}$ of the typical user to either a mmWave network or a $\mu$Wave network, respectively and the corresponding load on the associated cell. A general association metric is considered, in which a user is connected to a particular BS $x_i$ storing file $i$, with $k \in \{m, \mu\}$ if

$$x_i = \arg\max_{x_i\in\Phi_{k_i}} W_k r_{x_i}^{-\alpha_k}, \tag{8}$$

where $W_k$ is the association weight, the choice of which is based on $W_k = \mathrm{P}_k B_k$, where $B_k$ denotes the association bias corresponding to the network $k$ and the association is based on the

[4]From (2), perfect beam alignment yields the directivity gain $GG$.

[5]Using the probability transformation rule, the distribution of the power gain is converted to a normalized gamma distribution, namely a Nakagami distribution with the parameter $n_t^\mu - N_{x_i} + 1$. Furthermore, the effective interference channel power gain with cell load $N_t$, where $t \in \Phi_\mu\backslash\{x_i\}$, is also modelled as the gamma distribution, $H_t \sim \Gamma(N_t, 1)$ [30].

maximum biased received power. Accordingly, to calculate $p_{mw}$ and $p_{\mu w}$, we leverage the analysis from [31], and consider that the typical user is connected to the best network with respect to LTAB received power (*i.e.,* $B_m \mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_m}$ for mmWave network and $B_\mu \mathrm{P}_\mu r_{x_i}^{-\alpha_\mu}(n_t^\mu - N_{x_i} + 1)$ for $\mu$Wave network).

*Remark 2: The biased factor can be used to control the cell range and balance cell loads. By setting the biased factor to be greater than one, the cell coverage can be extended (to accommodate more users). Alternatively, the cell coverage can be shrunk by setting the biased factor to be smaller than one. Further, if the bias factor $B_k = 1$, the association is based on the maximum received power only.*

At this point, it is worth noting that unlike $\mu$Wave networks, it is important to characterize the least path loss distribution in mmWave networks by incorporating the effect of blockages. Accordingly, the least path loss distribution for a typical user in a mmWave network is given in Lemma 1, followed by the association probability of the typical user to $\mu$Wave network in Proposition 1.

**Lemma 1.** *The least path loss distribution for a typical user in a mmWave network is given as*

$$F_{\xi_{x_i}}^{\mathrm{mm}}(r) = 1 - \exp\Big( - \pi\lambda_m (r\mathrm{P}_m G_{x_i} B_m)^{\frac{2}{\alpha_{\mathcal{N}}}} - \tfrac{2\pi\lambda_m}{\beta^2}\big(1 - e^{-\beta(r\mathrm{P}_m G_{x_i} B_m)^{\frac{1}{\alpha_{\mathcal{L}}}}}\big(1 + \beta(r\mathrm{P}_m G_{x_i} B_m)^{\frac{1}{\alpha_{\mathcal{L}}}}\big)\big) + \tfrac{2\pi\lambda_m}{\beta^2}\big(1 - e^{-\beta(r\mathrm{P}_m G_{x_i} B_m)^{\frac{1}{\alpha_{\mathcal{N}}}}}\big(1 + \beta(r\mathrm{P}_m G_{x_i} B_m)^{\frac{1}{\alpha_{\mathcal{N}}}}\big)\big)\Big). \tag{9}$$

*Proof.* The proof of this Lemma follows from the proof of Proposition 1 of [32]. However, for convenience, we present a sketch of the proof here. Consider a point process, where the points represent the path loss between the typical user and randomly placed BSs in a mmWave network. Let $\phi_{\mathrm{mm}} = \left\{\xi_{x_i} \triangleq \frac{r_{x_i}^{\alpha_m}}{\mathrm{P}_m G_{x_i} B_m}\right\}$ be a homogeneous PPP of intensity $\lambda_m$. Here, the distance is a random variable, and its LOS state occurs with the probability of $e^{-\beta r}$. By using Mapping theorem [33, Theorem 2.34], the density function of this one dimensional PPP under the effect of blockages can be given as

$$\Lambda([0, r]) = \int_0^{(r\mathrm{P}_m G_{x_i} B_m)^{\frac{1}{\alpha_{\mathcal{L}}}}} 2\pi\lambda_m x e^{-\beta x} \mathrm{d}x + \int_0^{(r\mathrm{P}_m G_{x_i} B_m)^{\frac{1}{\alpha_{\mathcal{N}}}}} 2\pi\lambda_m x (1 - e^{-\beta x}) \mathrm{d}x. \tag{10}$$

Using the void probability of a PPP and with the help of (10), the least path loss distribution in a mmWave network can be given as (9).

□

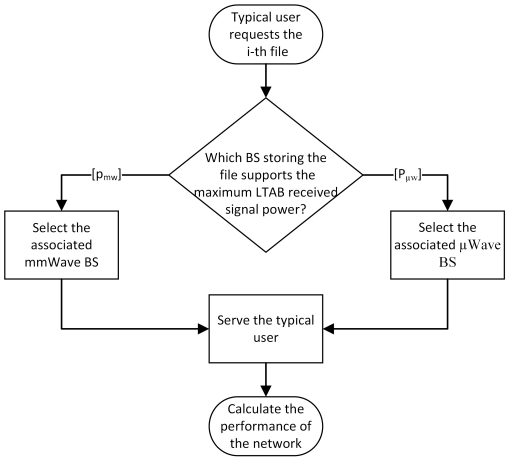


Fig. 2: An illustration of the communication policy for the typical user with either mmWave or $\mu$Wave network.

**Proposition 1.** *In a hybrid network consisting of mmWave and $\mu$Wave BSs as described, considering the least pass loss distribution, the probability that the typical user is connected to the $\mu$Wave network is given by*

$$p_{\mu w} = 2\pi\lambda_\mu \int_0^\infty r \exp\left(-\Lambda_m\left(\left(\frac{\bar{\mathrm{P}}_{\mathrm{mm}}}{\bar{\mathrm{P}}_\mu}\right)^{\frac{1}{\alpha_m}} r^{\frac{\alpha_\mu}{\alpha_m}}\right)\right) e^{-\pi\lambda_\mu r^2}\,\mathrm{d}r, \tag{11}$$

*where* $\bar{\mathrm{P}}_{\mathrm{mm}} = \mathrm{P}_m G_{x_i} B_m$, *with* $G_{x_i} = GG$ *and* $\bar{\mathrm{P}}_\mu = \mathrm{P}_\mu B_\mu n_t^\mu$, *with* $B_m = 2$, $B_\mu = 1$, *and*

$$\begin{aligned}\Lambda_m\left(\frac{\bar{\mathrm{P}}_{\mathrm{mm}}}{\bar{\mathrm{P}}_\mu}^{\frac{1}{\alpha_m}} r^{\frac{\alpha_\mu}{\alpha_m}}\right) =& \pi\lambda_m\left(\frac{\bar{\mathrm{P}}_{\mathrm{mm}}}{\bar{\mathrm{P}}_\mu}\right)^{\frac{1}{\alpha_\mathcal{N}}} r^{\frac{\alpha_\mu}{\alpha_\mathcal{N}}} - \frac{2\pi\lambda_m}{\beta^2}\left(1 - e^{-\beta\left(\frac{\bar{\mathrm{P}}_{\mathrm{mm}}}{\bar{\mathrm{P}}_\mu}\right)^{\frac{1}{\alpha_\mathcal{N}}} r^{\frac{\alpha_\mu}{\alpha_\mathcal{N}}}}\left(1 + \beta\left(\frac{\bar{\mathrm{P}}_{\mathrm{mm}}}{\bar{\mathrm{P}}_\mu}\right)^{\frac{1}{\alpha_\mathcal{N}}} r^{\frac{\alpha_\mu}{\alpha_\mathcal{N}}}\right)\right) \\ &+ \frac{2\pi\lambda_m}{\beta^2}\left(1 - e^{-\beta\left(\frac{\bar{\mathrm{P}}_{\mathrm{mm}}}{\bar{\mathrm{P}}_\mu}\right)^{\frac{1}{\alpha_\mathcal{L}}} r^{\frac{\alpha_\mu}{\alpha_\mathcal{L}}}}\left(1 + \beta\left(\frac{\bar{\mathrm{P}}_{\mathrm{mm}}}{\bar{\mathrm{P}}_\mu}\right)^{\frac{1}{\alpha_\mathcal{L}}} r^{\frac{\alpha_\mu}{\alpha_\mathcal{L}}}\right)\right).\end{aligned} \tag{12}$$

*Proof.* The proof can be obtained by leveraging results of Lemma 1. □

Accordingly, for the above association probability of the typical user to the $\mu$Wave network, the requested file $i$ is served by the associated $\mu$Wave BS that can support the downlink rate greater than the target bit rate. Otherwise, the typical user will be associated to the mmWave network with the probability $p_{mw} = 1 - p_{\mu w}$. After obtaining the association probabilities, we now give approximations for the cell load given in (6) and (7) in the following Lemma.

**Lemma 2.** *The approximated mean cell loads at the associated BS and other BSs, except the associated BS in the hybrid network can respectively be given as [34]*

$$\bar{N}_{x_i} = 1 + \frac{1.28\lambda_u p_{qw}}{\lambda_q}, \text{ and } \bar{N}_{\tilde{x}_i} = \frac{\lambda_u p_{qw}}{\lambda_q}, \tag{13}$$

*where $x_i \in \Phi_m \cup \Phi_\mu$ is the associated BS, and $q \in \{m, \mu\}$ depending on the context. Further, $\tilde{x}_i \in \Phi_m \cup \Phi_\mu \backslash \{x_i\}$ denotes all other BSs, except the associated BS.*

*Proof.* The proof of this lemma is given in Appendix A. □

Unless otherwise stated, hereinafter the above approximated cell load will be used throughout the paper. A pictorial representation summarizing the communication policy in the hybrid network is given in Fig. 2 at the top of previous page.

## V. Performance metric

To design the caching placement scheme, we use the ASP of file delivery as the performance metric, which is defined as the successful response to the user's request, which occurs when the downlink rate is more than the target bit rate of the file. Thus, when the typical user requests the $i$th file, the ASP of file delivery can be given as

$$\mathcal{P}_s(\{\nu_i\}) = \sum\nolimits_{i=1}^{L} f_i \, \mathbb{P}\left[\mathcal{R}_{x_i} \geq \nu_i\right], \tag{14}$$

where $f_i$ is the probability of requesting the $i$th file, $\nu_i$ is the normalized target bit rate of file $i$ and $\mathcal{R}_{x_i}$ is normalized supported data rate supported by the serving BS $x_i \in \Phi_{m_i} \cup \Phi_{\mu_i}$. Now, as the mmWave and $\mu$Wave networks follow two independent PPPs, it is possible to perform the analysis on both the processes independently with an association probability. In this regard, we assume that the typical user can communicate with all BSs that cache the requested file. Accordingly, the total ASP of file delivery can be given as

$$\mathcal{P}_s(\{\nu_i\}) = \mathcal{P}_s^{\mathrm{mm}}(\{\nu_i\}) p_{mw} + \mathcal{P}_s^{\mu}(\{\nu_i\}) p_{\mu w}, \tag{15}$$

where $\mathcal{P}_s^{\mathrm{mm}}(\{\nu_i\})$ and $\mathcal{P}_s^{\mu}(\{\nu_i\})$ denote the conditional ASP of file delivery by the mmWave and $\mu$Wave networks, respectively.

**Proposition 2.** *The ASP of file delivery by the mmWave network is tightly upper bounded by*

$$\begin{aligned}
\mathcal{P}_s^{mm}(\{\nu_i\}) \leq \sum\nolimits_{i=1}^{L} \Bigg\{ f_i \Bigg\{ \sum\nolimits_{l=1}^{\hat{m}} \binom{\hat{m}}{l} (-1)^{l+1} \sum\nolimits_{j \in \{\mathcal{L},\mathcal{N}\}} \Bigg\{ p_j \exp\Big( \frac{-AlQ_i \sigma_m^2}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}} \Big) \\
\times \Bigg\{ \prod_{\tilde{q}=1}^{3} \prod_{\tilde{j} \in \{\mathcal{L},\mathcal{N}\}} \exp\Big[ -\lambda_{m_i} \int_{r_{x_i}}^{\infty} \Big( 1 - \Big( \frac{1}{1 + \frac{AlQ_i \tilde{G}_{\tilde{q}} r^{-\alpha_{\tilde{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} \hat{m}}} \Big)^{\hat{m}} \Big) p_{\tilde{j}} 2\pi p_{\tilde{q}} \, r dr \Big] \Bigg\} \\
\times \Bigg\{ \prod_{\hat{q}=1}^{3} \prod_{\hat{j} \in \{\mathcal{L},\mathcal{N}\}} \exp\Big[ -\overline{\lambda}_{m_i} \int_{0}^{\infty} \Big( 1 - \Big( \frac{1}{1 + \frac{AlQ_i \hat{G}_{\hat{q}} r^{-\alpha_{\hat{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} \hat{m}}} \Big)^{\hat{m}} \Big) p_{\hat{j}} 2\pi p_{\hat{q}} \, r dr \Big] \Bigg\} \Bigg\} \Bigg\} \Bigg\} \Bigg\} \Bigg\} ,
\end{aligned} \tag{16}$$

*where* $G_{x_i} = GG$, $\tilde{G}_{\tilde{q}}, \hat{G}_{\hat{q}} \in \{\tilde{G}_1 = \hat{G}_1 = GG, \tilde{G}_2 = \hat{G}_2 = Gg, \tilde{G}_3 = \hat{G}_3 = gg\}$, $p_{\tilde{q}}$, *and* $p_{\hat{q}}$ *with* $\tilde{q}, \hat{q} \in \{1,2,3\}$ *are probabilities of related effective antenna gains. Particularly,* $p_j, p_{\tilde{j}}, p_{\hat{j}} \in \{p_{\mathcal{N}} = 1 - e^{-\beta r_{x_i}}, p_{\mathcal{L}} = e^{-\beta r_{x_i}}\}$ *with* $j, \tilde{j}, \hat{j} \in \{\mathcal{L}, \mathcal{N}\}$ *being the probabilities of the channel being LOS or NLOS. Further,* $A = \hat{m}(\hat{m}!)^{\frac{-1}{\hat{m}}}$, *where for mmWave network,* $\hat{m} > 1$, $Q_i = 2^{\bar{N}_{x_i} \nu_i} - 1$, *and* $\overline{\lambda}_{m_i} = (1 - p_{m_i}) \lambda_m$.

*Proof.* The proof of this proposition is given in Appendix B. □

**Proposition 3.** *The ASP of file delivery by the $\mu$Wave network is tightly upper bounded by*

$$\mathcal{P}_s^{\mu}(\{\nu_i\}) \leq \sum_{i=1}^{L} f_i \left\{ \sum_{l=1}^{n_t^{\mu}-\bar{N}_{x_i}+1} \left\{ \binom{n_t^{\mu}-\bar{N}_{x_i}+1}{l} (-1)^{l+1} \exp\left(\frac{-l\bar{A}Q_i\sigma_{\mu}^2}{\mathrm{P}_{\mu}(n_t^{\mu}-\bar{N}_{x_i}+1)r_{x_i}^{-\alpha_{\mu}}}\right) \right.\right.$$
$$\times \left\{ \exp\left[-2\pi\lambda_{\mu_i} \int_{r_{x_i}}^{\infty} r\left(1-\left(1+\frac{\bar{A}lQ_i r^{-\alpha_{\mu}}}{r_{x_i}^{-\alpha_{\mu}}(n_t^{\mu}-\bar{N}_{x_i}+1)}\right)^{-\bar{N}_{\bar{x}_i}}\right)dr\right] \right\}$$
$$\left.\left. \times \left\{ \exp\left[-2\pi\overline{\lambda}_{\mu_i} \int_{0}^{\infty} r\left(1-\left(1+\frac{\bar{A}lQ_i r^{-\alpha_{\mu}}}{r_{x_i}^{-\alpha_{\mu}}(n_t^{\mu}-\bar{N}_{x_i}+1)}\right)^{-\bar{N}_{\bar{x}_i}}\right)dr\right] \right\} \right\}, \tag{17}$$

*where* $Q_i$[6]$= 2^{\bar{N}_{x_i}\nu_i} - 1$, $\overline{\lambda}_{\mu_i} = \lambda_{\mu}(1-p_{\mu_i})$, *and* $\bar{A} = (n_t^{\mu}-\bar{N}_{x_i}+1)[(n_t^{\mu}-\bar{N}_{x_i}+1)!]^{-1/(n_t^{\mu}-\bar{N}_{x_i}+1)}$. *All other parameters are generally the same as those defined for the mmWave case but with notational changes.*

*Proof.* Due to page limitations, the proof of this proposition is omitted but can be obtained in a similar way as the proof of Proposition 2. □

## VI. Proposed Caching Placement in the Hybrid Network

To place the contents in the hybrid network, we aim to optimize the ASP of file delivery by considering a finite memory size and content popularity. This can be achieved by optimally determining the caching probabilities in the caching placement phase. Accordingly, we formulate an optimization problem as:

$$\mathbf{P1}: \max_{\{p_{m_i}\},\{p_{\mu_i}\}} \mathcal{P}_s(\{\nu_i\}), \tag{18}$$

$$\text{s.t.} \quad \sum_{i=1}^{L} p_{m_i} \leq \mathrm{C}_m, \tag{19}$$

$$\sum_{i=1}^{L} p_{\mu_i} \leq \mathrm{C}_{\mu}, \tag{20}$$

$$0 \leq p_{m_i} \leq 1 \;\; \& \;\; 0 \leq p_{\mu_i} \leq 1\,, \;\; \forall i \in \mathcal{S}. \tag{21}$$

The constraints in (19) and (20) ensure that the size of the total cached files should be less than or equal to $\mathrm{C}_m$ for mmWave and $\mathrm{C}_{\mu}$ for $\mu$Wave networks. In the above, the ASP of file delivery is given as in (22), shown on the top of next page. The functions consisting of variables $p_{m_i}$ and $p_{\mu_i}$ are exponential functions, which are convex. Hence, their summation is also convex. However, due to the binomial term $(-1)^{l+1}$, the objective function is no longer convex. In particular, it is a *DC* function and hence a DC problem (DCP), which makes it quite rigorous to cope with. In this regard, we consider the noise-limited (NL) and interference-limited (IL) scenarios separately to try and simplify the problem. Accordingly, we reformulate the optimization problem under two different scenarios.

[6]By normalizing the whole bandwidth, the maximum normalized data rate is set as 1 and the target data rate $\nu_i \in [0,1]$. For simplicity, we consider the same normalized target rate for both mmWave and $\mu$Wave networks.

$$
\begin{aligned}
&\mathcal{P}_s(\{\nu_i\}) = \mathcal{P}_s^{\mathrm{mm}}(\{\nu_i\})p_{mw} + \mathcal{P}_s^{\mu}(\{\nu_i\})p_{\mu w} \\
&\leq \sum_{i=1}^{L} \Bigg\{ f_i \Bigg\{ \sum_{l=1}^{\hat{m}} \binom{\hat{m}}{l} (-1)^{l+1} \sum_{j\in\{\mathcal{L},\mathcal{N}\}} \Bigg\{ p_j \exp\Big( \frac{-AlQ_i\sigma_m^2}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}} \Big) \\
&\times \Big\{ \exp\Big[ -\lambda_{m_i} \underbrace{\sum_{\tilde{q}=1}^{3} \sum_{\tilde{j}\in\{\mathcal{L},\mathcal{N}\}} \int_{r_{x_i}}^{\infty} \Big( 1 - \Big( \frac{1}{1 + \frac{AlQ_i \tilde{G}_{\tilde{q}} r^{-\alpha_{\tilde{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} \hat{m}}} \Big)^{\hat{m}} \Big) p_{\tilde{j}} 2\pi p_{\tilde{q}}\, r \mathrm{d}r}_{Z'(i,l)} \Big] \Big\} \\
&\times \Bigg\{ \exp\Big[ -\overline{\lambda}_{m_i} \underbrace{\sum_{\hat{q}=1}^{3} \sum_{\hat{j}\in\{\mathcal{L},\mathcal{N}\}} \int_{0}^{\infty} \Big( 1 - \Big( \frac{1}{1 + \frac{AlQ_i \hat{G}_{\hat{q}} r^{-\alpha_{\hat{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} m}} \Big)^{\hat{m}} \Big) p_{\hat{j}} 2\pi p_{\hat{q}}\, r dr}_{Z''(i,l)} \Big] \Bigg\} \Bigg\} \Bigg\} \Bigg\} \Bigg\} p_{mw} \\
&+ \sum_{i=1}^{L} f_i \Bigg\{ \sum_{l=1}^{n_t^{\mu} - \bar{N}_{x_i} + 1} \Bigg\{ \binom{n_t^{\mu} - \bar{N}_{x_i} + 1}{l} (-1)^{l+1} \exp\Big( \frac{-l\bar{A}Q_i\sigma_{\mu}^2}{\mathrm{P}_{\mu}(n_t^{\mu} - \bar{N}_{x_i} + 1) r_{x_i}^{-\alpha_{\mu}}} \Big) \\
&\times \Big\{ \exp[ -\lambda_{\mu_i} \underbrace{\int_{r_{x_i}}^{\infty} 2\pi r \Big( 1 - \Big( 1 + \frac{\bar{A}lQ_i r^{-\alpha_{\mu}}}{r_{x_i}^{-\alpha_{\mu}} (n_t^{\mu} - \bar{N}_{x_i} + 1)} \Big)^{-\bar{N}_{\tilde{x}_i}} \Big) dr}_{W'(i,l)} ] \Big\} \\
&\times \Bigg\{ \exp[ -\overline{\lambda}_{\mu_i} \underbrace{\int_{0}^{\infty} 2\pi r \Big( 1 - \Big( 1 + \frac{\bar{A}lQ_i r^{-\alpha_{\mu}}}{r_{x_i}^{-\alpha_{\mu}} (n_t^{\mu} - \bar{N}_{x_i} + 1)} \Big)^{-\bar{N}_{\tilde{x}_i}} \Big) dr}_{W''(i,l)} ] \Bigg\} \Bigg\} p_{\mu w}
\end{aligned} \tag{22}
$$

### *A. Optimal caching probabilities under NL scenario*

Recent studies on mmWave networks [29], state that mmWave networks in urban settings are more NL than IL. This is due to the fact that in the presence of blockages, the signals received from unintentional sources are close to negligible. In such densely blocked scenarios (typical for urban settings), signal to noise ratio (SNR) provides a good enough approximation to SINR for directional mmWave networks. Furthermore, for the $\mu$Wave case, when the number of users is much smaller than the number of serving BSs, the impact of interference is negligible when compared to the noise power [17]. Accordingly, the typical user can be served by sharing limited resources with other users. Hence, in this section we ignore the interference part $B$ in (5) and (7), and consider the NL case only to analyze the effects of channel selection diversity on optimal caching strategy. Additionally, such an assumption also aids us in deriving closed form expressions for the ASP of file delivery.

**Proposition 4.** *The ASP of file delivery by the mmWave network under a NL scenario is given*

*by*

$$\mathcal{P}_s^{mm}(\{\nu_i\}) = \sum\nolimits_{i=1}^{L} f_i \left[1 - \exp\Big( - \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} k_j p_{m_i} Z_j(\frac{\eta_i G_{x_i}}{Q_i}) - \hat{k} p_{m_i} (\frac{\eta_i G_{x_i}}{Q_i})^{\delta_{m_\mathcal{N}}} \Big)\right] \tag{23}$$

*where* $k_j = c_j \frac{\pi \lambda_m \hat{m}^{\hat{m}} \delta_{m_j}}{\Gamma(\hat{m})}$, $c_j \in \{-1, 1\}$, *with* $j \in \{\mathcal{L}, \mathcal{N}\}$, *such that* $c_\mathcal{L} = 1, c_\mathcal{N} = -1$, $\hat{k} = \pi \lambda_m \frac{\Gamma(\delta_{m_\mathcal{N}} + \hat{m})}{\hat{m}^{\delta_{m_\mathcal{N}}} \Gamma(\hat{m})}$, $\delta_{m_j} = \frac{2}{\alpha_j}$, $\eta_i = \frac{\mathrm{P_m}}{\sigma_m^2}$, $Q_i = 2^{\bar{N}_{x_i}\nu_i} - 1$. *Further,* $Z_j(\tilde{\omega})$ *is defined as*

$$Z_j(\tilde{\omega}) = \int_0^\infty \int_0^{\tilde{\omega}} \frac{exp(-\frac{\hat{m}}{\omega}\psi)}{\omega^{(\hat{m}+1)}} d\omega \psi^{(\delta_{m_j} + \hat{m} - 1)} exp(-\beta \psi^{\frac{\delta_{m_j}}{2}}) d\psi \,,$$

*where* $\tilde{\omega} = \frac{\eta_i G_{x_i}}{Q_i}$.

*Proof.* The proof of this proposition is given in Appendix C. □

Similarly, the ASP of file delivery in the $\mu$Wave network can be derived with the channel power gain now being gamma distributed, where blockage effects and beamforming gains are ignored.

**Proposition 5.** *The ASP of file delivery in a $\mu$Wave network under a NL scenario is given by*

$$\mathcal{P}_s^{\mu}(\{\nu_i\}) = \sum\nolimits_{i=1}^{L} f_i \Big\{ 1 - \exp(-\tilde{k}\ p_{\mu_i} (\frac{\tilde{\eta}_i}{Q_i(n_t^\mu - \bar{N}_{x_i} + 1)})^{\delta_\mu}) \Big\}, \tag{24}$$

*where* $\tilde{k} = \frac{\pi \lambda_\mu \Gamma(\delta_\mu + n_t^\mu - \bar{N}_{x_i} + 1)}{(n_t^\mu - \bar{N}_{x_i} + 1)^{\delta_\mu} \Gamma(n_t^\mu - \bar{N}_{x_i} + 1)}$, $\tilde{\eta}_i = \frac{\mathrm{P}_\mu}{\sigma_\mu^2}$, $\delta_\mu = \frac{2}{\alpha_\mu}$, $Q_i = 2^{\bar{N}_{x_i}\nu_i} - 1$.

*Proof.* The proof of this proposition follows in a similar way as the proof of Proposition 4. □

Now, the optimization problem **P1** in (18)–(21) is convex and the objective function can be rewritten as

$$\begin{aligned}\mathcal{P}_s(\{\nu_i\}) = &\sum\nolimits_{i=1}^{L} f_i \left[1 - \exp\Big( - \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} k_j p_{m_i} Z_j(\frac{\eta_i G_{x_i}}{Q_i}) - \hat{k} p_{m_i} (\frac{\eta_i G_{x_i}}{Q_i})^{\delta_{m_\mathcal{N}}} \Big)\right] p_{mw} \\ &+ \sum\nolimits_{i=1}^{L} f_i \left[1 - \exp(-\tilde{k}\ p_{\mu_i} (\frac{\tilde{\eta}_i}{Q_i(n_t^\mu - \bar{N}_{x_i} + 1)})^{\delta_\mu})\right] p_{\mu w}. \end{aligned} \tag{25}$$

Accordingly, the Lagrangian function of the reformulated problem can be given as

$$\begin{aligned}&\mathcal{L}\left(\{p_{m_i}\}, \{p_{\mu_i}\}, \tilde{\omega}, \hat{\omega}, \{\tilde{\mu}_i\}, \{\hat{\mu}_i\}\right) \\ &= \sum\nolimits_{i=1}^{L} f_i \left[1 - \exp\Big( - \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} k_j p_{m_i} Z_j(\frac{\eta_i G_{x_i}}{Q_i}) - \hat{k} p_{m_i} (\frac{\eta_i G_{x_i}}{Q_i})^{\delta_{m_\mathcal{N}}} \Big)\right] p_{mw} \\ &+ \sum\nolimits_{i=1}^{L} f_i \Big\{ 1 - \exp(-\tilde{k}\ p_{\mu_i} (\frac{\tilde{\eta}_i}{Q_i(n_t^\mu - \bar{N}_{x_i} + 1)})^{\delta_\mu}) \Big\}\ p_{\mu w} - \tilde{\omega} \left( \sum\nolimits_{i=1}^{L} p_{m_i} - \mathrm{C}_m \right) - \hat{\omega} \left( \sum\nolimits_{i=1}^{L} p_{\mu_i} - \mathrm{C}_\mu \right) \\ &- \sum\nolimits_{i=1}^{L} \tilde{\mu}_i (p_{m_i} - 1) - \sum\nolimits_{i=1}^{L} \hat{\mu}_i (p_{\mu_i} - 1), \end{aligned} \tag{26}$$

where $\tilde{\omega}, \hat{\omega}, \tilde{\mu}_i$ and $\hat{\mu}_i$ are the Lagrangian multipliers associated with the constraints (19)-(20), respectively. Since Slater's condition is satisfied for the problem, its optimal solution can be obtained by solving its dual, which can be written as

$$\min_{\tilde{\omega}, \hat{\omega}, \{\tilde{\mu}_i\}, \{\hat{\mu}_i\} \geq 0} \ \max_{\{p_{m_i}\}, \{p_{\mu_i}\}} \mathcal{L}\left(\{p_{m_i}\}, \{p_{\mu_i}\}, \tilde{\omega}, \hat{\omega}, \{\tilde{\mu}_i\}, \{\hat{\mu}_i\}\right). \tag{27}$$

**Algorithm 1 : Computation of $\{p^*_{m_i}\}$ and $\{p^*_{\mu_i}\}$**

1: Initialize : $\tilde{\omega}, \hat{\omega}, \{\tilde{\mu}_i\}, \{\hat{\mu}_i\}$

2: Compute $p_{m_i}$ using (29).

3: Compute $p_{\mu_i}$ using (30).

4: Update the Langrangian multipliers $\tilde{\omega}, \hat{\omega}, \{\tilde{\mu}_i\}$, and $\{\hat{\mu}_i\}$.

5: Repeat steps $2-4$ until convergence.

6: $p^*_{m_i} \leftarrow p_{m_i}(\tilde{\omega}^*, \tilde{\mu}^*_i), p^*_{\mu_i} \leftarrow p_{\mu_i}(\hat{\omega}^*, \hat{\mu}^*_i)$.

The dual problem in (27) is then solved in an iterative fashion which alternates between a *sub-problem* of updating the caching probability variables $\{p_{m_i}\}$ and $\{p_{\mu_i}\}$ by fixing the Lagrangian multipliers $(\tilde{\omega}, \hat{\omega}, \{\tilde{\mu}_i\}, \{\hat{\mu}_i\})$, and a *master problem* of computing new Lagrangian multipliers based on the obtained caching probabilities. Further, in the sub-problem, by taking the partial derivative of (26) with respect to $p_{m_i}$, we can find the optimal caching probabilities for mmWave BSs. Accordingly,

$$\frac{\partial \mathcal{L}(\{p_{m_i}\}, \tilde{\omega}, \{\tilde{\mu}_i\})}{\partial p_{m_i}} = f_i \exp[p_{m_i}(-\tilde{A}_i - \tilde{B}_i)] + \tilde{\omega} + \tilde{\mu}_i \tag{28}$$

where $\tilde{A}_i = \sum_{j\in\{\mathcal{L},\mathcal{N}\}} k_j Z_j (\frac{\eta_i G_{x_i}}{Q_i})$ and $\tilde{B}_i = \hat{k}(\frac{\eta_i G_{x_i}}{Q_i})^{\delta_{m_\mathcal{N}}}$. The optimal caching probability is now given by

$$p_{m_i} = \left[\frac{1}{\tilde{A}_i + \tilde{B}_i} \log\left(\frac{f_i p_{mw}}{\tilde{\omega} + \tilde{\mu}_i}\right)\right]^+, \tag{29}$$

where $[x]^+ = \max\{0, x\}$. Likewise, the optimal caching probability for $\mu$Wave BS is given by

$$p_{\mu_i} = \left[\frac{1}{\hat{T}_i \tilde{k}} \log\left(\frac{p_{\mu w} f_i}{\hat{\omega} + \hat{\mu}_i}\right)\right]^+, \tag{30}$$

where $\hat{T}_i = (\frac{\tilde{\eta}_i}{Q_i(n_t^\mu - N_{x_i}+1)})^{\delta_\mu}$. The Lagrange multipliers are updated using subgradient method. The algorithm to find optimal caching probabilities $\{p^*_{m_i}\}$ and $\{p^*_{\mu_i}\}$ is given in Algorithm 1.

### *B. Optimal caching probabilities under IL scenario*

Compared to the NL case, the interference part in (5) and (7) will be dominant when the number of mm/$\mu$Wave BSs (*i.e.,* density) increases and blockage density in the network decreases. In this subsection, the ASP of file delivery under a IL case is given.

**Proposition 6.** *The IL ASP of file delivery in a mmWave network is given by*

$$\mathcal{P}_s^{mm}(\nu_i) \leq \sum_{i=1}^{L} \left\{ f_i \left\{ \sum_{l=1}^{\hat{m}} \binom{\hat{m}}{l} (-1)^{l+1} \sum_{j\in\{\mathcal{L},\mathcal{N}\}} p_j \right.\right.$$
$$\left.\left. \times \left\{ \prod_{\tilde{q}=1}^{3} \prod_{\tilde{j}\in\{\mathcal{L},\mathcal{N}\}} \exp\Big[ -\lambda_{m_i} \int_{r_{x_i}}^{\infty} \Big(1 - \Big(\frac{1}{1 + \frac{AlQ_i \tilde{G}_{\tilde{q}} r^{-\alpha_{\tilde{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} \hat{m}}}\Big)^{\hat{m}}\Big) p_{\tilde{j}} 2\pi p_{\tilde{q}}\, r dr \Big] \right\} \right\}$$

$$\times \left\{ \prod_{\hat{q}=1}^{3} \prod_{\hat{j}\in\{\mathcal{L},\mathcal{N}\}} \exp\Big[ -\overline{\lambda}_{m_i} \int_0^{\infty} \Big(1 - \Big(\frac{1}{1+\frac{AlQ_i\hat{G}_{\hat{q}} r^{-\alpha_{\hat{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} \hat{m}}}\Big)^{\hat{m}}\Big) p_{\hat{j}} 2\pi p_{\hat{q}}\, r dr \Big] \right\} \right\} \right\}. \tag{31}$$

*Proof.* This proof can be obtained similar to Proposition 2 with interference part only. □

**Proposition 7.** *The IL ASP of file delivery in a $\mu$Wave network is given similarly as*

$$\mathcal{P}_s^{\mu}(\nu_i) \leq \sum_{i=1}^{L} f_i \left\{ \sum_{l=1}^{n_t^{\mu}-\bar{N}_{x_i}+1} \left\{ \binom{n_t^{\mu}-\bar{N}_{x_i}+1}{l} (-1)^{l+1} \right.\right.$$
$$\left\{ \exp\Big[ -2\pi\lambda_{\mu_i} \int_{r_{x_i}}^{\infty} r \Big(1 - \big(1 + \frac{l\bar{A}Q_i r^{-\alpha_\mu}}{r_{x_i}^{-\alpha_\mu}(n_t^{\mu}-\bar{N}_{x_i}+1)}\big)^{-\bar{N}_{\bar{x}_i}}\Big) dr \Big] \right\}$$
$$\left. \left. \left\{ \exp\Big[ -2\pi\overline{\lambda}_{\mu_i} \int_0^{\infty} r \Big(1 - \big(1 + \frac{l\bar{A}Q_i r^{-\alpha_\mu}}{r_{x_i}^{-\alpha_\mu}(n_t^{\mu}-\bar{N}_{x_i}+1)}\big)^{-\bar{N}_{\bar{x}_i}}\Big) dr \Big] \right\} \right\} \right\}, \tag{32}$$

*where all the parameters are as defined before.*

*Proof.* This proof can be obtained by leveraging the derivation of Proposition 3. □

After establishing the ASP of file delivery for the hybrid network in an IL scenario, the objective function of the optimization problem **P1** for this case can now be rewritten as in (33), shown on the top of the next page. Note that, $h(\{p_{m_i}\}), g(\{p_{m_i}\}), hh(\{p_{\mu_i}\}), gg(\{p_{\mu_i}\})$ in (33) are convex. Hence, the optimization problem to find the optimal file placement scheme under IL scenario can be formulated as a standard *DCP problem*, given as

$$\mathbf{P2}: \quad \min_{\{p_{m_i}\},\{p_{\mu_i}\}} \; [-h(\{p_{m_i}\}) + g(\{p_{m_i}\}) - hh(\{p_{\mu_i}\}) + gg(\{p_{\mu_i}\})] \tag{34}$$

$$\text{s.t.} \quad \sum_{i=1}^{L} p_{m_i} \leq \mathrm{C}_m, \tag{35}$$

$$\sum_{i=1}^{L} p_{\mu_i} \leq \mathrm{C}_{\mu}, \tag{36}$$

$$0 \leq p_{m_i} \leq 1, \;\; \& \;\; 0 \leq p_{\mu_i} \leq 1\,, \quad \forall i \in \mathcal{S}. \tag{37}$$

Like before, since mmWave and $\mu$Wave networks can be independently represented through their respective association probabilities, we can separately calculate the optimal caching probabilities for mmWave BSs and $\mu$Wave BSs by rewriting the above optimization problem into two sub-DCP problems. Based on [35], we propose an iterative algorithm as given in Algorithm 2 to obtain the optimal caching probabilities for both mmWave and $\mu$Wave networks by separately converting the sub-DC objective functions to convex functions. Below, we provide the proof of convergence of Algorithm 2.

**Proposition 8.** *The original objective function in* (34) *of the DCP problem **P2** can be convexified by replacing it with its upper bound. The proposed algorithm is then convergent with respect to an increase in the iteration number.*

$$
\begin{aligned}
&\mathcal{P}_s(\{\nu_i\}) \le \sum\nolimits_{i=1}^{L} \Bigg\{ f_i \Bigg\{ \sum\nolimits_{l=1}^{\hat{m}} \binom{\hat{m}}{l} (-1)^{l+1} \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} p_j \\
&\Big\{ \prod_{\tilde{q}=1}^{3} \prod_{\tilde{j}\in\{\mathcal{L},\mathcal{N}\}} \exp\Big[ -\lambda_{m_i} \int_{r_{x_i}}^{\infty} \Big( 1 - \Big( \frac{1}{1 + \frac{AlQ_i \tilde{G}_{\tilde{q}} r^{-\alpha_{\tilde{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} \hat{m}}} \Big)^{\hat{m}} \Big) p_{\tilde{j}} 2\pi p_{\tilde{q}}\, r dr \Big] \Big\} \\
&\Bigg\{ \prod_{\hat{q}=1}^{3} \prod_{\hat{j}\in\{\mathcal{L},\mathcal{N}\}} \exp\Big[ -\overline{\lambda}_{m_i} \int_{0}^{\infty} \Big( 1 - \Big( \frac{1}{1 + \frac{AlQ_i \hat{G}_{\hat{q}} r^{-\alpha_{\hat{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} \hat{m}}} \Big)^{\hat{m}} \Big) p_{\hat{j}} 2\pi p_{\hat{q}}\, r dr \Big] \Bigg\} \Bigg\} \Bigg\} \; p_{mw} \\
&+ \sum\nolimits_{i=1}^{L} f_i \Bigg\{ \sum_{l=1}^{n_t^{\mu} - \bar{N}_{x_i} + 1} \Bigg\{ \binom{n_t^{\mu} - \bar{N}_{x_i} + 1}{l} (-1)^{l+1} \exp\Big[ -2\pi\lambda_{\mu_i} \int_{r_{x_i}}^{\infty} r \Big( 1 - \big( 1 + \frac{l\bar{A}Q_i r^{-\alpha_\mu}}{r_{x_i}^{-\alpha_\mu} (n_t^{\mu} - \bar{N}_{x_i} + 1)} \big)^{-\bar{N}_{\tilde{x}_i}} \Big) dr \Big] \\
&\exp\Big[ -2\pi\overline{\lambda}_{\mu_i} \int_{0}^{\infty} r \Big( 1 - \big( 1 + \frac{l\bar{A}Q_i r^{-\alpha_\mu}}{r_{x_i}^{-\alpha_\mu} (n_t^{\mu} - \bar{N}_{x_i} + 1)} \big)^{-\bar{N}_{\tilde{x}_i}} \Big) dr \Big] \Bigg\} \Bigg\} \; p_{\mu w} \\
&= \sum\nolimits_{i=1}^{L} f_i \sum\nolimits_{l=1}^{\hat{m}} \binom{\hat{m}}{l} (-1)^{(l+1)} \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} p_j \\
&\times \exp\Bigg[ p_{m_i} \lambda_m \underbrace{\sum_{\tilde{q}=1}^{3} \sum\nolimits_{\tilde{j}\in\{\mathcal{L},\mathcal{N}\}} \int_{0}^{r_{x_i}} \Big( 1 - \Big( \frac{1}{1 + \frac{AlQ_i \tilde{G}_{\tilde{q}} r^{-\alpha_{\tilde{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} \hat{m}}} \Big)^{\hat{m}} \Big) p_{\tilde{j}} 2\pi p_{\tilde{q}}\, r dr}_{\tilde{Z}(i,l)} \Bigg] \\
&\times \exp\Bigg[ -\lambda_m \underbrace{\sum_{\hat{q}=1}^{3} \sum\nolimits_{\hat{j}\in\{\mathcal{L},\mathcal{N}\}} \int_{0}^{\infty} \Big( 1 - \Big( \frac{1}{1 + \frac{AlQ_i \hat{G}_{\hat{q}} r^{-\alpha_{\hat{j}}}}{G_{x_i} r_{x_i}^{-\alpha_j} \hat{m}}} \Big)^{\hat{m}} \Big) p_{\hat{j}} 2\pi p_{\hat{q}}\, r_t dr_t}_{\hat{Z}(i,l)} \Bigg] p_{mw} \\
&+ \sum_{i=1}^{L} f_i \sum_{l=1}^{n_t^{\mu} - \bar{N}_{x_i} + 1} \binom{n_t^{\mu} - \bar{N}_{x_i} + 1}{l} (-1)^{(l+1)} \times \exp\Big[ p_{\mu_i} \lambda_\mu \underbrace{\int_{0}^{r_{x_i}} r \Big( 1 - \big( 1 + \frac{l\bar{A}Q_i r^{-\alpha_\mu}}{r_{x_i}^{-\alpha_\mu} (n_t^{\mu} - \bar{N}_{x_i} + 1)} \big)^{-\bar{N}_{\tilde{x}_i}} \Big) 2\pi dr}_{\bar{W}(i,l)} \Big] \\
&\times \exp\Big[ -\lambda_\mu \underbrace{\int_{0}^{\infty} r \Big( 1 - \big( 1 + \frac{l\bar{A}Q_i r^{-\alpha_\mu}}{r_{x_i}^{-\alpha_\mu} (n_t^{\mu} - \bar{N}_{x_i} + 1)} \big)^{-\bar{N}_{\tilde{x}_i}} \Big) 2\pi dr}_{\hat{W}(i,l)} \Big] p_{\mu w} \\
&= \underbrace{\sum\nolimits_{i=1}^{L} f_i \sum\nolimits_{l=\text{odd number}} \binom{\hat{m}}{l} \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} p_j \exp[-\lambda_m \hat{Z}(i,l)] \exp[p_{m_i} \lambda_m \tilde{Z}(i,l)] p_{mw}}_{h(\{p_{m_i}\})} \\
&- \underbrace{\sum\nolimits_{i=1}^{L} f_i \sum\nolimits_{l=\text{even number}} \binom{\hat{m}}{l} \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} p_j \exp[-\lambda_m \hat{Z}(i,l)] \exp[p_{m_i} \lambda_m \tilde{Z}(i,l)] p_{mw}}_{g(\{p_m\})} \\
&+ \underbrace{\sum\nolimits_{i=1}^{L} f_i \sum\nolimits_{l=\text{odd number}} \binom{n_t^{\mu} - \bar{N}_{x_i} + 1}{l} \exp[-\lambda_\mu \hat{W}(i,l)] \exp[p_{\mu_i} \lambda_\mu \tilde{W}(i,l)] p_{\mu w}}_{hh(\{p_{\mu_i}\})} \\
&- \underbrace{\sum\nolimits_{i=1}^{L} f_i \sum\nolimits_{l=\text{even number}} \binom{n_t^{\mu} - \bar{N}_{x_i} + 1}{l} \exp[-\lambda_\mu \hat{W}(i,l)] \exp[p_{\mu_i} \lambda_\mu \tilde{W}(i,l)] p_{\mu w}}_{gg(\{p_{\mu_i}\})}
\end{aligned} \tag{33}
$$

**Algorithm 2 : Computation of $\{p^*_{m_i}\}$ and $\{p^*_{\mu_i}\}$**

**Initialize** : counter $k = 0, \{p^0_{m_i}\}, \{p^0_{\mu_i}\},$ step size $\Delta = 10^{-4}$ and threshold $\hat{\delta} = 10^{-5}$

**Repeat**

**Compute:** $\hat{h}(\mathbf{p}_m^{k+1}; \mathbf{p}_m^k) = h(\mathbf{p}_m^k) + \nabla \mathbf{h}^{\mathrm{T}}(\mathbf{p}_m^k)(\mathbf{p}_m^{k+1} - \mathbf{p}_m^k)$

$\hat{hh}(\mathbf{p}_\mu^{k+1}; \mathbf{p}_\mu^k) = hh(\mathbf{p}_\mu^k) + \nabla \mathbf{hh}^{\mathrm{T}}(\mathbf{p}_\mu^k)\,(\mathbf{p}_\mu^{k+1} - \mathbf{p}_\mu^k)$

**Solve 1:** Set the value of $\{p^{k+1}_{m_i}\}$ to be a solution of

minimize $g(\{p^{k+1}_{m_i}\}) - \hat{h}(\{p^{k+1}_{m_i}\}; \{p^k_{m_i}\})$

subject to $\sum_{i=1}^{L} p^{k+1}_{m_i} \leq C_m,\ 0 \leq p^{k+1}_{m_i} \leq 1,\ \forall i \in \mathcal{S}$

**Solve 2:** Set the value of $\{p^{k+1}_{\mu_i}\}$ to be a solution of

minimize $gg(\{p^{k+1}_{\mu_i}\}) - \hat{hh}(\{p^{k+1}_{\mu_i}\}; \{p^k_{\mu_i}\})$

subject to $\sum_{i=1}^{L} p^{k+1}_{\mu_i} \leq C_\mu,\ 0 \leq p^{k+1}_{\mu_i} \leq 1,\ \forall i \in \mathcal{S}$

**Update:** $k = k + 1$.

$\mathbf{p}_m^k = \mathbf{p}_m^{k-1},$

$\mathbf{p}_\mu^k = \mathbf{p}_\mu^{k-1},$

**Until** convergence or maximum iteration number is reached.

*Proof.* Let $\mathbf{p}_m^k$ and $\mathbf{p}_\mu^k$ be the feasible points for problem **P2**. Applying Taylor series approximation on $h(\{p^{k+1}_{m_i}\})$ and $hh(\{p^{k+1}_{\mu_i}\})$ at feasible points $\mathbf{p}_m^k$ and $\mathbf{p}_\mu^k$, the objective function is rewritten as

$$v_{k+1} = g(\mathbf{p}_m^{k+1}) - h(\mathbf{p}_m^{k+1}) + gg(\mathbf{p}_\mu^{k+1}) - hh(\mathbf{p}_\mu^{k+1}), \tag{38}$$

such that (38) is now convex. Since the region of feasible solution remains the same, the feasible points $\mathbf{p}_m^k$ and $\mathbf{p}_\mu^k$ are also feasible for the convexified problem and other feasible points $\mathbf{p}_m^{k+1}$ and $\mathbf{p}_\mu^{k+1}$ that exist for the convexified problem are also the feasible points of the problem **P2**. Furthermore, for all $\mathbf{p}_m$ and $\mathbf{p}_\mu$ using Taylor series approximation, the convexity of $h$ and $hh$ gives us

$$\hat{h}(\mathbf{p}_m^{k+1}; \mathbf{p}_m^k) \leq h(\mathbf{p}_m^{k+1}),\ \text{and}\ \hat{hh}(\mathbf{p}_\mu^{k+1}; \mathbf{p}_\mu^k) \leq hh(\mathbf{p}_\mu^{k+1}). \tag{39}$$

Hence, if $\mathbf{p}_m^0$ and $\mathbf{p}_\mu^0$ are chosen to be feasible, all corresponding iterates will be feasible. Now, we show that the objective value converges over the iterations. According to the inequalities

$$g(\mathbf{p}_m) < h(\mathbf{p}_m),\ gg(\mathbf{p}_\mu) < hh(\mathbf{p}_\mu),\ \text{and}\ v_{k+1} \leq \underbrace{g(\mathbf{p}_m^{k+1}) - \hat{h}(\mathbf{p}_m^{k+1}; \mathbf{p}_m^k) + gg(\mathbf{p}_\mu^{k+1}) - \hat{hh}(\mathbf{p}_\mu^{k+1}; \mathbf{p}_\mu^k)}_{\hat{v}_{k+1}},$$

we minimize the value of $\hat{v}_{k+1}$ at each iteration $k$, and obtain $v_k$ by using previous $\mathbf{p}_m^k$ and $\mathbf{p}_\mu^k$ such that

$$v_k \geq \hat{v}_{k+1} \geq v_{k+1}. \tag{40}$$

The value of the above objective function is now non-increasing and will always converge, possibly to negative infinity, which concludes the proof. □

TABLE I: Parameter values

| **Parameter notation** | **Physical meaning** | **Values** |
|---|---|---|
| $\eta_i = \frac{\mathrm{P}_m}{\sigma^2_{m_{x_i}}}, \forall i \in \mathcal{S}$ | SNR of the typical user for file $i$ from the mmWave serving BS | 54 (dB) |
| $\tilde{\eta}_i = \frac{\mathrm{P}_\mu}{\sigma^2_{\mu_{x_i}}}, \forall i \in \mathcal{S}$ | SNR of the typical user for file $i$ from the $\mu$Wave serving BS | 104 (dB) |
| $\theta$ | Mainlobe beamwidth | $\pi/6$ |
| $G^M$ / $G^m$ | Mainlobe antenna gain / sidelobe antenna gain | 10 (dB) / -10 (dB) |
| $G_{x_i}$ | Effective antenna gain between the serving mmWave BS and typical user | 100 (dB) |
| $\alpha_{\mathcal{L}}$ / $\alpha_{\mathcal{N}}$ | Path loss exponent of LoS and NLoS | 2 / 4 |
| $\alpha_\mu$ | $\mu$Wave path loss exponent | 3.5 |
| $\lambda_m$ | mmWave BS density | $5\times10^{-6}$ (nodes/m$^2$) |
| $\lambda_\mu$ | $\mu$Wave BS density | $10^{-6}$ (nodes/m$^2$) |
| $\beta$ | Blockage density | 0.01 |
| $\upsilon$ | Skewness of the content popularity | 0.8 |
| $L$ | The number of files | 10 |
| $\mathrm{C}_m/\mathrm{C}_\mu$ | Cache size of mmWave/$\mu$Wave BS | 5 / 6 |
| $\hat{m}$ | Nakagami fading parameter for mmWave channel | 3 |
| $n_t^\mu$ | The number of $\mu$Wave antennas | 8 |
| $\nu_{\max}$ | The maximum normalized file delivery rate | 1 |
| $\nu_i, \forall i \in \mathcal{S}$ | The normalized rate for $i$-th file delivery (*i.e.,*$\nu_i \in [0, \nu_{\max}]$) | 0.08 |
| $B_m$ / $B_\mu$ | Bias factors | 2 / 1 |

## VII. Numerical results

After developing the analytical framework in the previous sections, we now evaluate the performance of the proposed caching placement strategy with respect to Algorithm 1 and 2. Unless otherwise stated, most of the parameters used and their corresponding values are inspired from literature and given in Table I. For simplicity, a uniform target rate for each file is considered throughout the analysis.

We begin by evaluating the optimal caching probabilities in a NL hybrid network using Algorithm 1 for varying densities of mmWave and $\mu$Wave networks and different Nakagami parameters in Fig. 3. It is worth noting that according to the CDF of the process $\{\frac{r^\alpha}{\mathcal{X}}\}$ in (C.8), when $\lambda_m p_{m_i}$ becomes higher, the minimum of the CDF of $\{\frac{r^\alpha}{\mathcal{X}}\}$ increases. Accordingly, the ASP of file delivery will be higher due to the increase in intensity measure. In fact, the reciprocal of $\{\frac{r^\alpha}{\mathcal{X}}\}$ represents the effective channel gain and hence, with the increase in density, the probability to obtain higher channel gain also increases. However, it can also be seen that when the density increases, the caching probabilities for the most popular files decrease and tend to be uniformly distributed. This means that by sacrificing higher channel gain for a few specific contents, we can increase the hitting probability of all contents (*content diversity gain*) such that the optimal ASP of file delivery can be achieved. Therefore, there is a tradeoff between channel gain and cache hit. Besides, in the same figure, we also evaluate the effect of Nakagami fading parameter $\hat{m}$, which relates to the channel power gain. It can be seen that the proposed algorithm is

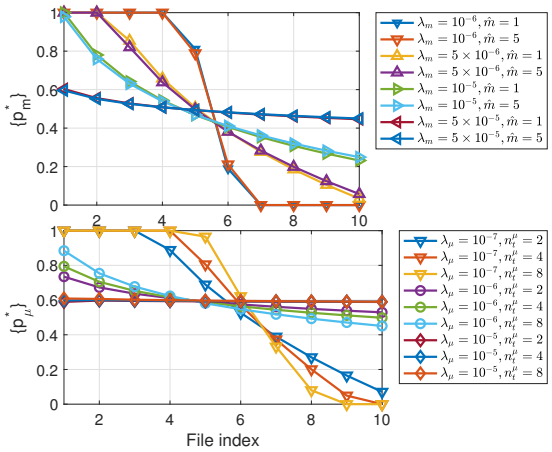


Fig. 3: Optimal caching probability v.s. varying $\lambda$ and $\hat{m}$ values under NL scenario.

minimally affected by $\hat{m}$. For example, for a particular mmWave BS density $\lambda_m$, the gap in performance of the optimal caching probabilities between Nakagami parameters $\hat{m} = 1$ and $5$ is unsubstantial. This result provides the design insight that $\lambda_m$ is a more decisive parameter than $\hat{m}$ when designing optimal caching schemes for hybrid networks. The above explanation holds true for $\mu$Wave systems as well as can be seen from the tradeoff in performance for $\mu$Wave optimal caching probabilities in terms of BS density.

To obtain further insights, we now evaluate the performance of the proposed content placement strategy for the NL case in terms of ASP of file delivery.

*1) Comparison with other common proactive caching schemes:* For the sake of comparison, we consider three different content placement schemes: 1) caching $M$ most popular contents (MC), 2) caching contents uniformly (UC), and 3) caching contents randomly (RC). It is evident from Fig. 4 that the proposed caching scheme is superior to the MC, UC and RC for varying content popularities. When the skewness $\upsilon$ of the content popularity distribution is close to zero, the meaning of which is that the content popularity is uniformly distributed and uniformly requested by users, the proposed caching placement strategy is significantly better than the others except UC. On the other hand, while MC is better than RC for higher values of $\upsilon$ and vice versa for low $\upsilon$, UC performs better than RC throughout the entire range of $\upsilon$.

*2) Effect of blockages:* It can be seen from Fig. 5 that when blockage density increases, the ASP of file delivery decreases. This is due to the fact that increasing blockages in the mmWave network results in the attenuation of the received signal. However, the decrease in optimal ASP is not very significant. This can be explained as: i) for a substantial blockage density, with

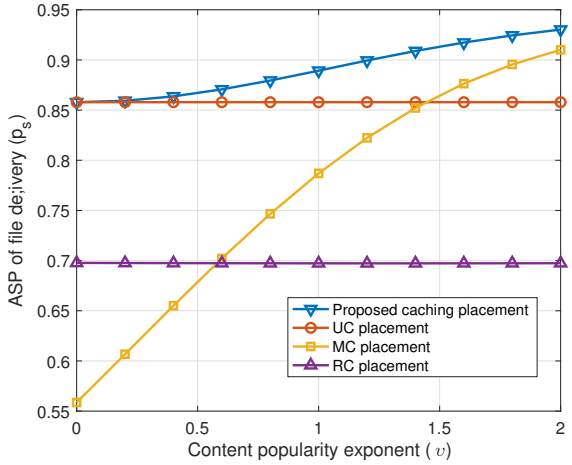


Fig. 4: ASP of file delivery for various caching placement

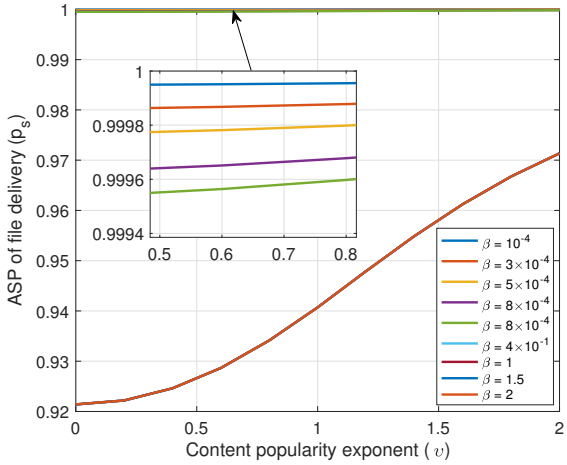


Fig. 5: ASP of file delivery for various blockage densities

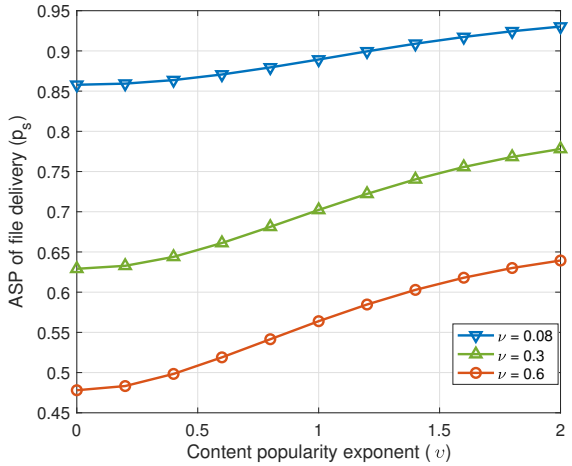


Fig. 6: ASP of file delivery for various QoS thresholds under NL scenario.

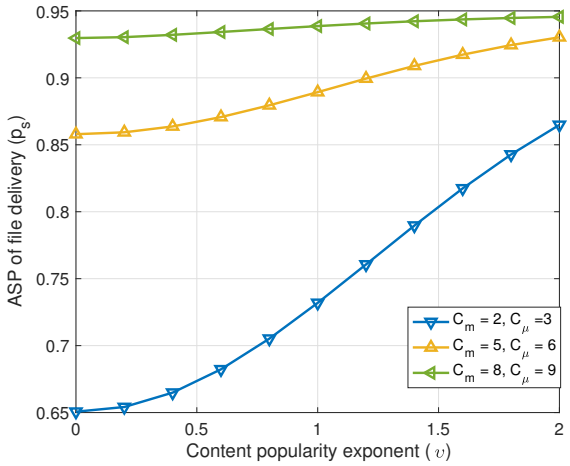


Fig. 7: ASP of file delivery for various cache sizes under NL scenario.

the increase in skewness, the number of files with higher probability requests decreases, ii) the proposed algorithm makes sure that the higher caching probability of the most requested files is stored in the $\mu$Wave BSs, which are not affected by blockages, and iii) for the mmWave network, the algorithm also makes sure that the most requested files are stored only in the mmWave BSs with less average probability of NLOS than LOS. Hence, it can be concluded that the proposed algorithm is a *blockage-aware* optimal caching strategy.

*3) Effect of QoS requirements:* It can be seen from Fig. 6 that as the target QoS (normalized target data rate) increases, the performance of the system decreases for fixed cache sizes of $C_m = 5$ and $C_\mu = 6$. This is due to the fact that increasing the QoS requirements, increases the threshold of the average success probability. Due to propagation factors such as blockages, path

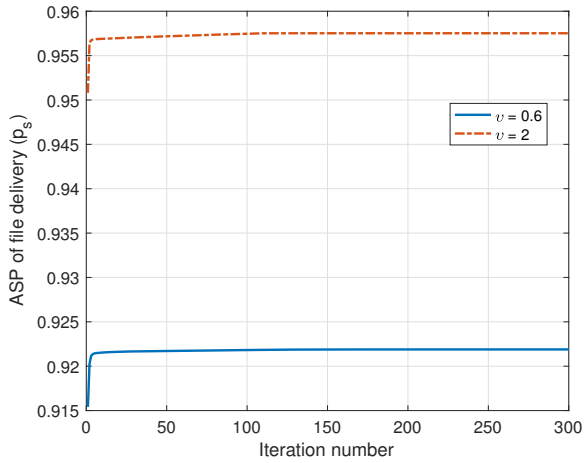


Fig. 8: Convergence of Algorithm 2 under IL scenario.

loss, etc., several BSs in the network may no longer support the QoS requirements, thus failing to meet the association criteria even though they might have the requested file.

*4) Effect of cache size:* The effect of QoS requirements can somewhat be nullified by increasing the cache size of either mmWave or $\mu$Wave BSs as can be seen from Fig. 7. Here, target data rate $\nu = 0.08$. When the cache size increases, the ASP of file delivery performance also improves. This is due to the fact that with a larger cache size, the probability of storing a particular file is higher, which directly relates to the ASP of file delivery.

After validating the results for the NL scenario, we now do the same for the IL case with $\lambda_\mu = 10^{-5}$, $\lambda_m = 5 \times 10^{-5}$, and $\beta = 0.005$. We begin by showing the evolution of Algorithm 2 in Fig. 8 for $\upsilon = \{0.6, 2\}$. The monotonic increase of the cost function (ASP of file delivery) verifies the convergence of the proposed algorithm, which was also proved in Proposition 8.

Similar to the NL case, we now evaluate the performance of the proposed content placement strategy as given in Algorithm 2 for the IL scenario.

*1) Comparison with other common proactive caching schemes:* It can be seen from Fig. 9 that like the NL case, the proposed caching placement scheme is also mostly superior to the MC, UC and RC in the IL case for varying content popularities. When the skewness $\upsilon$ of content popularity distribution is close to zero, the proposed caching placement strategy is distinctively better than the others. However, at higher values of $\upsilon$, performance of MC is comparable to or slightly better than the proposed algorithm. This is due to the fact that higher values of $\upsilon$ means that the content popularity is not uniform and the contents are not uniformly requested by the users. Instead, only the most popular files are requested by the users which gives a performance

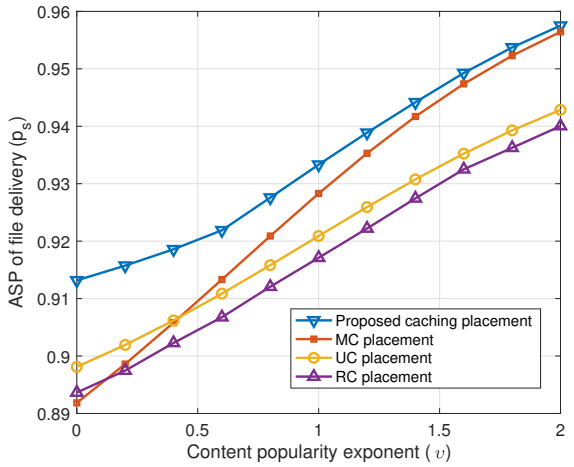


Fig. 9: ASP of file delivery for various caching placement

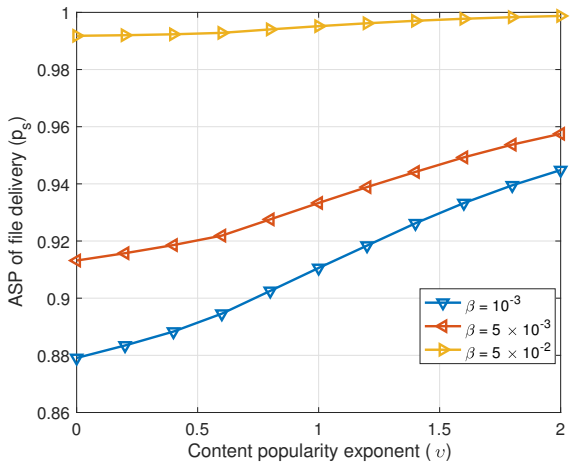


Fig. 10: ASP of file delivery for various blockage densities

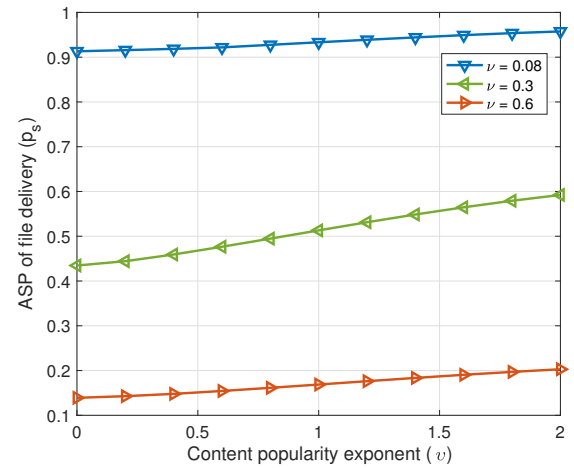


Fig. 11: ASP of file delivery for various QoS requirements under IL scenario.

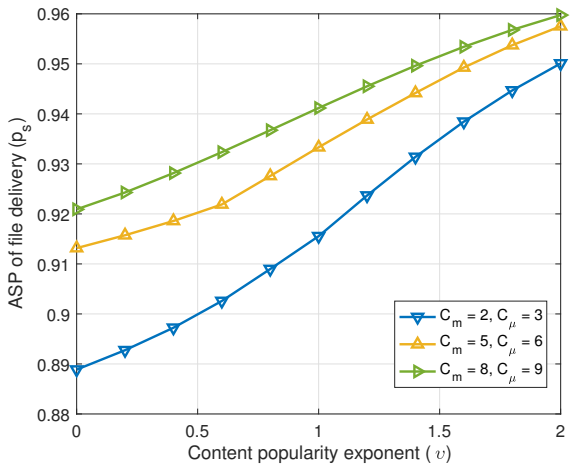


Fig. 12: ASP of file delivery for various cache sizes under IL scenario.

edge to MC. Furthermore, while MC is better than RC for higher values of $\upsilon$ and close to RC for low $\upsilon$, UC perform better than RC throughout the entire range of $\upsilon$.

*2) Effect of blockages:* Unlike the NL case, when blockage density increases, the ASP of file delivery increases in the IL case as can be seen from Fig. 10. The increase in optimal ASP is due to the effect of interference mitigation through blockages. More blockages in the network help in attenuating the interfering signals, which enhances the rate of file delivery. This can be considered as one of the very few instances when blockages are beneficial.

*3) Effect of QoS requirements and cache size:* The effects of QoS requirements and cache size on the ASP of file delivery follow similar trends in the IL case as for the NL case. These can be seen from Fig. 11 and Fig. 12, respectively.

## VIII. Conclusion

A framework to study the optimal probabilistic caching strategy at desirable BSs in a mm-$\mu$Wave hybrid network was presented. To obtain design insights, essential factors such as interference among BSs, blockages in the mmWave network, uncertainties both in node locations and channel fading, path loss, and loads at BSs, were incorporated in the system model. In particular, by considering the ASP of file delivery as the performance metric, the content placement strategy was formulated as an optimization problem and accordingly, two algorithms were provided, one each for noise-limited and interference-limited scenarios to acquire the optimal caching probabilities. Detailed numerical analysis was performed to evaluate the performance of the proposed optimal caching placement strategy with respect to several essential parameters, which demonstrated the superiority of the proposed scheme over others, albeit certain trade-offs.

## Appendix A
## Proof of Lemma 2

By assuming that the distribution of the association area of the serving BS is proportional to that of a traditional distance-based Poisson-Voronoi (PV) process with the same mean area, the approximated distribution of association area for the mmWave and $\mu$Wave BSs can be written as functions of $\frac{\lambda_m}{p_{mw}}$ and $\frac{\lambda_\mu}{p_{\mu w}}$, respectively, which are denoted by $C_m(\frac{\lambda_m}{p_{mw}})$ and $C_\mu(\frac{\lambda_\mu}{p_{\mu w}})$. Now, the biased association area distribution (*i.e.,* a weighted PV) of $C_q^{'}$, under the assumption as made in [34] is proportional to its of association area and can be written as $f_{C_q^{'}}(c) \propto c f_{C_q}(c)$. Then, the biased association distribution can be given as [34]

$$f_{C_q^{'}}(c) = \frac{c f_{C_q}(c)}{\mathbb{E}[C_q]} = \frac{3.5^{3.5}}{\Gamma(3.5)} A(Ac)^{3.5} \exp(-3.5Ac)\,. \tag{A.1}$$

where $A = \frac{\lambda_m}{p_{mw}}$ for $q = m$ and $A = \frac{\lambda_\mu}{p_{\mu w}}$ for $q = \mu$. At this point, it is worth noting that the moments of a Poisson random variable, $X \sim \text{Pois}(\lambda)$ with density $\lambda$ can be written in terms of Stirling numbers of the second kind $S(n,k)$ [34]. Accordingly, $\mathbb{E}[X^n] = \sum_{k=0}^{n} \lambda^k S(n,k)$. Hence, by a slight abuse of notation, the mean of the number of other users (except the typical user), that are associated with the serving BS $x_i \in \Phi_{m_i} \cup \Phi_{\mu_i}$ storing the file $i$ can be given as

$$\mathbb{E}[N_{x_i}^n] = \mathbb{E}_{C_q^{'}}\left[\mathbb{E}\left[N_{x_i}^n | C_q^{'}\right]\right] = \sum\nolimits_{k=1}^{n} \lambda_u^k S(n,k) \mathbb{E}[C_q^{'k}]\,. \tag{A.2}$$

In the above, $q = m$ gives $C_m^{'}$, whose distribution is linked to $\frac{\lambda_m}{p_{mw}}$ and $q = \mu$ gives $C_\mu^{'}$, whose distribution is linked to $\frac{\lambda_\mu}{p_{\mu w}}$. In the above, the number of users $N \in \{1, \cdots, n\}$. Now, using (A.1) we can acquire the approximation of the association area as

$$\mathbb{E}[N_{x_i}^n] = \sum\nolimits_{k=1}^{n} (\lambda_u A)^k S(n,k) \mathbb{E}[C_q^{k+1}(1)]\,. \tag{A.3}$$

Accordingly, the first moment of the load (including the typical user) on the associated BS is obtained by setting $n=1$, and can be given as

$$\bar{N}_{x_i} = 1 + \mathbb{E}[N_{x_i}] \overset{(a)}{=} 1 + 1.28\frac{\lambda_u p_{qw}}{\lambda_q}, \tag{A.4}$$

where $(a)$ in the above is obtained from [36]. Similarly, the average cell load on any other BS, except for the associated BS can be approximated as $\bar{N}_{\tilde{x}_i} = \frac{\lambda_u p_{qw}}{\lambda_q}$, which concludes the proof.

## APPENDIX B
## PROOF OF PROPOSITION 2

To proof this proposition, we substitute (5) and (6) into (14). Accordingly, the conditional ASP of file delivery by the mmWave BSs can be reduced to

$$\begin{aligned}
\mathcal{P}_s^{\mathrm{mm}}(\nu_i) &= \sum\nolimits_{i=1}^{L} f_i\, \mathbb{P}[\gamma_{m_{x_i}} \geq \underbrace{2^{\bar{N}_{x_i}\nu_i} - 1}_{Q_i}] \\
&= \sum\nolimits_{i=1}^{L} f_i\, \mathbb{P}\left[ \frac{\mathrm{P}_m G_{x_i} \mathcal{X}_{x_i} r_{x_i}^{-\alpha_j}}{\left( \sigma_m^2 + \underbrace{\sum\limits_{t\in\Phi^c_{m_i}} \mathrm{P}_m G_t \mathcal{X}_t r_t^{-\alpha_{\hat{j}}}}_{I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}}} + \underbrace{\sum\limits_{\substack{k\in\Phi_{m_i}\\ k\neq x_i}} \mathbf{P}_m G_k \mathcal{X}_k r_k^{-\alpha_{\tilde{j}}}}_{I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}} \right)} \geq Q_i \right] \\
&\overset{(a)}{\leq} \sum\nolimits_{i=1}^{L} f_i\, \mathbb{E}_{\alpha_j,\, I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}},\, I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}} \left\{ 1 - \left[ 1 - \exp\left( \frac{-AQ_i(\sigma_m^2 + I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}} + I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}})}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}} \right) \right]^{\hat{m}} \right\} \\
&\overset{(b)}{=} \sum\nolimits_{i=1}^{L} f_i\, \mathbb{E}_{\alpha_j,\, I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}},\, I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}} \left\{ 1 - \sum\nolimits_{l=0}^{\hat{m}} \binom{\hat{m}}{l} (-1)^l \exp\left( \frac{-AlQ_i(\sigma_m^2 + I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}} + I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}})}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}} \right) \right\} \\
&\overset{(c)}{=} \sum\nolimits_{i=1}^{L} f_i \sum\nolimits_{l=1}^{\hat{m}} \binom{\hat{m}}{l} (-1)^{l+1} \mathbb{E}_{I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}}, I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}} \left\{ \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} p_j \exp\left( \frac{-AlQ_i(\sigma_m^2 + I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}} + I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}})}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}} \right) \right\} \\
&\overset{(d)}{=} \sum\nolimits_{i=1}^{L} f_i \sum\nolimits_{l=1}^{\hat{m}} \binom{\hat{m}}{l} (-1)^{l+1} \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} p_j \mathbb{E}_{I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}}, I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}} \left\{ \exp\left( \frac{-AlQ_i(\sigma_m^2 + I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}} + I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}})}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}} \right) \right\} \\
&= \sum\nolimits_{i=1}^{L} f_i \sum\nolimits_{l=1}^{\hat{m}} \binom{\hat{m}}{l} (-1)^{l+1} \sum_{j\in\{\mathcal{L},\mathcal{N}\}} p_j \exp\left( \frac{-AlQ_i\sigma_m^2}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}} \right) \mathbb{E}_{I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}}} \left\{ \exp\left( \frac{-AlQ_i I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}}}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}} \right) \right\} \\
&\qquad \times \mathbb{E}_{I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}} \left\{ \exp\left( \frac{-AlQ_i I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}} \right) \right\}.
\end{aligned} \tag{B.1}$$

Here, $x_i$ refers to the associated mmWave BS that has the requested file $i$ in its local cache. Particularly, $(a)$ follows from the tight lower bound of a gamma random variable [28], and by taking unconditional expectation with respect to path loss exponent $\alpha_j$ and interference $I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}}$ and $I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}$. Further, $(b)$ follows from the Binomial theorem, $(c)$ is obtained by taking average over $\alpha_j$, and $(d)$ is obtained by exploiting the independence of $p_j$ over $I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}}$ and $I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}$. Now, applying the thinning theorem of a PPP by considering blockages and effective antenna gains, the point process $I^{\tilde{G},\tilde{j}}_{\Phi_{m_i}}$ and $I^{\hat{G},\hat{j}}_{\Phi^c_{m_i}}$ can be divided into 6 independent sub-PPPs as shown in the following, respectively.

$$I_{\Phi_{m_i}}^{\tilde{G},\tilde{j}} = I_{\Phi_{m_i}}^{GG,\mathcal{L}} + I_{\Phi_{m_i}}^{Gg,\mathcal{L}} + I_{\Phi_{m_i}}^{gg,\mathcal{L}} + I_{\Phi_{mi}}^{GG,\mathcal{N}} + I_{\Phi_{m_i}}^{Gg,\mathcal{N}} + I_{\Phi_{m_i}}^{gg,\mathcal{N}}, \tag{B.2}$$

$$I_{\Phi_{m_i}^c}^{\hat{G},\hat{j}} = I_{\Phi_{m_i}^c}^{GG,\mathcal{L}} + I_{\Phi_{m_i}^c}^{Gg,\mathcal{L}} + I_{\Phi_{m_i}^c}^{gg,\mathcal{L}} + I_{\Phi_{mi}^c}^{GG,\mathcal{N}} + I_{\Phi_{m_i}^c}^{Gg,\mathcal{N}} + I_{\Phi_{m_i}^c}^{gg,\mathcal{N}}. \tag{B.3}$$

Accordingly, the expectation part in (B.1) can be reduced to

$$\mathbb{E}_{I_{\Phi_{m_i}}^{\tilde{G}_{\tilde{q}},\tilde{j}}}\left\{\exp\left(\frac{-AlQ_i(I_{\Phi_{m_i}}^{GG,\mathcal{L}} + I_{\Phi_{m_i}}^{Gg,\mathcal{L}} + I_{\Phi_{m_i}}^{gg,\mathcal{L}} + I_{\Phi_{mi}}^{GG,\mathcal{N}} + I_{\Phi_{m_i}}^{Gg,\mathcal{N}} + I_{\Phi_{m_i}}^{gg,\mathcal{N}})}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_{\mathcal{L}}}}\right)\right\}$$
$$= \prod_{\tilde{q}=1}^{3}\prod_{\tilde{j}\in\{\mathcal{L},\mathcal{N}\}} \mathbb{E}_{I_{\Phi_{m_i}}^{\tilde{G}_{\tilde{q}},\tilde{j}}}\left\{\exp\left(\frac{-AlQ_i I_{\Phi_{m_i}}^{\tilde{G}_{\tilde{q}},\tilde{j}}}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}}\right)\right\}, \tag{B.4}$$

$$\mathbb{E}_{I_{\Phi_{m_i}^c}^{\hat{G}_{\hat{q}},\hat{j}}}\left\{\exp\left(\frac{-AlQ_i(I_{\Phi_{m_i}^c}^{GG,\mathcal{L}} + I_{\Phi_{m_i}^c}^{Gg,\mathcal{L}} + I_{\Phi_{m_i}^c}^{gg,\mathcal{L}} + I_{\Phi_{mi}^c}^{GG,\mathcal{N}} + I_{\Phi_{m_i}^c}^{Gg,\mathcal{N}} + I_{\Phi_{m_i}^c}^{gg,\mathcal{N}})}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_{\mathcal{L}}}}\right)\right\}$$
$$= \prod_{\hat{q}=1}^{3}\prod_{\hat{j}\in\{\mathcal{L},\mathcal{N}\}} \mathbb{E}_{I_{\Phi_{m_i}^c}^{\hat{G}_{\hat{q}},\hat{j}}}\left\{\exp\left(\frac{-AlQ_i I_{\Phi_{m_i}}^{\hat{G}_{\hat{q}},\hat{j}}}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}}\right)\right\}, \tag{B.5}$$

where (B.4) and (B.5) follows from the fact that sub-PPPs in (B.2) and (B.3) are independent, respectively. Below we compute the expectation of $I_{\Phi_{m_i}}^{GG,\mathcal{L}}$ and $I_{\Phi_{m_i}^c}^{GG,\mathcal{L}}$ only. All other terms in (B.4) and (B.5) can be derived in a similar way. By utilizing Laplace transform, the above expectation can be given as

$$\mathbb{E}\left[\exp\left(-s_{(i,l,j)} I_{\Phi_{m_i}}^{\tilde{G}_1,\tilde{j}}\right)\middle|\tilde{G}_1 = GG, \tilde{j} = \mathcal{L}\right] \overset{(f)}{=} \mathbb{E}_{\Phi_{m_i}^{\tilde{G}_1,\mathcal{L}}}\Big\{\prod_{k\in\Phi_{m_i}^{\tilde{G}_1,\mathcal{L}}} \mathbb{E}_{\mathcal{X}_k}\left[\exp(-s_{(i,l,j)}\mathrm{P}_m GG\mathcal{X}_k r^{-\alpha_{\mathcal{L}}})\right]\Big\}$$
$$\overset{(g)}{=} \mathbb{E}_{\Phi_{m_i}^{\tilde{G}_1,\mathcal{L}}}\Big\{\prod_{k\in\Phi_{m_i}^{\tilde{G}_1,\mathcal{L}}}\Big(\frac{1}{1+\frac{s_{(i,l,j)}\mathrm{P}_m GG r^{-\alpha_{\mathcal{L}}}}{\hat{m}}}\Big)^{\hat{m}}\Big\}$$
$$\overset{(h)}{=} \exp\Big[-\lambda_{m_i}\int_{r_{x_i}}^{\infty}\Big(1-\Big(\frac{1}{1+\frac{s_{(i,l,j)}\mathrm{P}_m GG r^{-\alpha_{\mathcal{L}}}}{\hat{m}}}\Big)^{\hat{m}}\Big)e^{-\beta r}2\pi p_1 r dr\Big], \tag{B.6}$$

$$\mathbb{E}\left[\exp\left(-s_{(i,l,j)} I_{\Phi_{m_i}^c}^{\hat{G}_1,\hat{j}}\right)\middle|\hat{G}_1 = GG, \hat{j} = \mathcal{L}\right] \overset{(i)}{=} \mathbb{E}_{\overline{\Phi}_{m_i}^{\hat{G}_1,\mathcal{L}}}\Big\{\prod_{t\in\overline{\Phi}_{m_i}^{\hat{G}_1,\mathcal{L}}} \mathbb{E}_{\mathcal{X}_t}\left[\exp(-s_{(i,l,j)}\mathrm{P}_m GG\mathcal{X}_t r^{-\alpha_{\mathcal{L}}})\right]\Big\}$$
$$\overset{(j)}{=} \mathbb{E}_{\overline{\Phi}_{m_i}^{\hat{G}_1,\mathcal{L}}}\Big\{\prod_{t\in\overline{\Phi}_{m_i}^{\hat{G}_1,\mathcal{L}}}\Big(\frac{1}{1+\frac{s_{(i,l,j)}\mathrm{P}_m GG r^{-\alpha_{\mathcal{L}}}}{\hat{m}}}\Big)^{\hat{m}}\Big\}$$
$$\overset{(k)}{=} \exp\Big[-\overline{\lambda}_{m_i}\int_{0}^{\infty}\Big(1-\Big(\frac{1}{1+\frac{s_{(i,l,j)}\mathrm{P}_m GG r^{-\alpha_{\mathcal{L}}}}{\hat{m}}}\Big)^{\hat{m}}\Big)e^{-\beta r}2\pi p_1 r dr\Big], \tag{B.7}$$

where $\Phi_{m_i}^c$ and $\overline{\Phi}_{m_i}$ have the same meaning but the use of the later is for notational simplicity. Further, $I_{\Phi_{m_i}}^{\tilde{G}_1,\tilde{j}}$ and $\overline{\Phi}_{m_i}^{\hat{G}_1,\hat{j}}$ are the point processes marked by the availability and unavailability of file $i$, $GG$ and LoS state, respectively. $s_{(i,l,j)} = \frac{AlQ_i}{\mathrm{P}_m G_{x_i} r_{x_i}^{-\alpha_j}}\Big|_{G_{x_i}=GG} = \frac{AlQ_i}{\mathrm{P}_m (GG) r_{x_i}^{-\alpha_j}}$ and $p_1$ is the probability when the antenna gain takes value $GG$ and $e^{-\beta r}$ is the probability of the occurrence of LoS transmission. In the above, $(f)$ and $(i)$ follow from independent channel fading gains, $(g)$ and $(j)$ follow from the moment generating function of a Nakagami-$\hat{m}$ random variable, and $(h)$ and $(k)$ follow from the probability generating functional of a PPP [33].

Now, taking into account all the sub-PPPs given by (B.2) and (B.3), we have

$$\mathbb{E}[\exp(-s_{(i,l,j)}I^{\tilde{G},\tilde{j}}_{\Phi^{\mathcal{L}}_{m_i}})] = \mathbb{E}[\exp(\sum\nolimits_{\tilde{q}=1}^{3}\sum\nolimits_{\tilde{j}\in\{L,N\}}(-s_{(i,l,j)}I^{\tilde{G}_{\tilde{q}},\tilde{j}}_{\Phi^{\mathcal{L}}_{m_i}}))] \overset{(l)}{=} \prod_{\tilde{q}=1}^{3}\prod_{\tilde{j}\in\{\mathcal{L},\mathcal{N}\}}\mathbb{E}[\exp(-s_{(i,l,j)}I^{\tilde{G}_{\tilde{q}},\tilde{j}}_{\Phi^{\mathcal{L}}_{m_i}})]$$

$$= \prod_{\tilde{q}=1}^{3}\prod_{\tilde{j}\in\{\mathcal{L},\mathcal{N}\}}\exp\Big[-\lambda_{m_i}\int_{r_{x_i}}^{\infty}\Big(1-\Big(\frac{1}{1+\frac{s_{(i,l,j)}\mathrm{P}_m\hat{G}_{\tilde{q}}r^{-\alpha_{\tilde{j}}}}{\tilde{m}}}\Big)^{\tilde{m}}\Big)p_{\tilde{j}}2\pi p_{\tilde{q}}r\mathrm{d}r\Big], \tag{B.8}$$

$$\mathbb{E}[\exp(-s_{(i,l,j)}I^{\hat{G},\hat{j}}_{\Phi^{\mathcal{L}}_{m_i}})] = \mathbb{E}[\exp(\sum\nolimits_{\hat{q}=1}^{3}\sum\nolimits_{\hat{j}\in\{L,N\}}(-s_{(i,l,j)}I^{\hat{G}_{\hat{q}},\hat{j}}_{\Phi^{\mathcal{L}}_{m_i}}))] \overset{(n)}{=} \prod_{\hat{q}=1}^{3}\prod_{\hat{j}\in\{\mathcal{L},\mathcal{N}\}}\mathbb{E}[\exp(-s_{(i,l,j)}I^{\hat{G}_{\hat{q}},\hat{j}}_{\Phi^{\mathcal{L}}_{m_i}})]$$

$$= \prod_{\hat{q}=1}^{3}\prod_{\hat{j}\in\{\mathcal{L},\mathcal{N}\}}\exp\Big[-\overline{\lambda}_{m_i}\int_{0}^{\infty}\Big(1-\Big(\frac{1}{1+\frac{s_{(i,l,j)}\mathrm{P}_m\hat{G}_{\hat{q}}r^{-\alpha_{\hat{j}}}}{\hat{m}}}\Big)^{\hat{m}}\Big)p_{\hat{j}}2\pi p_{\hat{q}}r\mathrm{d}r\Big], \tag{B.9}$$

where $(l)$ and $(n)$ follow from the fact that all sub-PPPs are independent. Finally, substituting the above results into (B.1), the desired proof is obtained.

## APPENDIX C
## PROOF OF PROPOSITION 4

Considering the NL scenario, the received SNR at the typical user in the mmWave network served by BS $x_i$ can be reduced to a form given as

$$\gamma^{\mathrm{SNR}}_{m_{x_i}} \approx \left(\mathrm{P}_m\, G_{x_i}\, \mathcal{X}_{x_i}\, r_{x_i}^{-\alpha_j}\right) / \left(\sigma_m^2\right).$$

Accordingly, the rate supported by the serving BS delivering file $i$ to the typical user is

$$R^{\mathrm{SNR}}_{m_{x_i}} \approx \frac{1}{\bar{N}_{x_i}}\log(1+\gamma^{\mathrm{SNR}}_{m_{x_i}}).$$

Now, the conditional probability of file delivery in the mmWave network is given by

$$\mathcal{P}_s^{mm}(\{\nu_i\}) = \sum\nolimits_{i=1}^{L} f_i\mathbb{P}[\log(1+\gamma^{\mathrm{SNR}}_{m_{x_i}}) \geq \bar{N}_{x_i}\nu_i]$$

$$= \sum\nolimits_{i=1}^{L} f_i\mathbb{P}[\frac{\mathrm{P}_m G_{x_i}\mathcal{X}_{x_i}r_{x_i}^{-\alpha_j}}{\sigma_m^2} \geq \underbrace{2^{\bar{N}_{x_i}\nu_i}-1}_{Q_i}] = \sum\nolimits_{i=1}^{L} f_i\mathbb{P}[\frac{r_{x_i}^{\alpha_j}}{\mathcal{X}_{x_i}} \leq \frac{\mathrm{P}_m G_{x_i}}{Q_i\sigma_m^2}], \tag{C.1}$$

where $\alpha_j = \{\alpha_{\mathcal{L}}, \alpha_{\mathcal{N}}\}$. According to thinning theorem, the PPP $\Phi_{m_i}$ is thinned from the process $\Phi_m$ with density $\lambda_m p_{m_i}$, where $p_{m_i}$ is the probability that the $i$th file is stored in the cache. Further, the process $\Phi_m$ consists of two sub-PPPs (*i.e.,* $\Phi^{\mathcal{L}}_{m_i}$ and $\Phi^{\mathcal{N}}_{m_i}$). Initially, it is necessary to calculate the density of $\Psi_i = \{r_{x_i}^{\alpha_j}(\triangleq ||\psi_q||)\} = \{||y_{i,q}||^{\alpha_j}\}$, where $i \in \mathcal{S}, q \in \mathbb{N}, j \in \{\mathcal{L},\mathcal{N}\}$. By using the mapping theorem, the intensity measure of the process $\Psi_i$ is given by

$$\Psi_i([0,\psi]) = \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}}\int_0^{\psi^{\frac{1}{\alpha_j}}}\lambda_m p_j p_{m_i}2\pi r\mathrm{d}r\,, \tag{C.2}$$

$$= \sum\nolimits_{j\in\{L,N\}} c_j 2\pi\lambda_m p_{m_i}(-\frac{1}{\beta})\Big\{\psi^{\frac{1}{\alpha_j}}\exp(-\beta\psi^{\frac{1}{\alpha_j}}) + \frac{1}{\beta}(\exp(-\beta\psi^{\frac{1}{\alpha_j}})-1)\Big\} + \lambda_m p_{m_i}\pi\psi^{\frac{2}{\alpha_{\mathcal{N}}}},$$

where $c_j \in \{-1,1\}, c_{\mathcal{L}} = 1, c_{\mathcal{N}} = -1$. Then the density is given by

$$\lambda_{\Psi_i}(\psi) = \frac{\mathrm{d}\Psi_i([0,\psi])}{\mathrm{d}\psi}$$

$$\begin{aligned}
&\overset{(a)}{=} \sum_{j\in\{\mathcal{L},\mathcal{N}\}} \Big\{ c_j\lambda_m p_{m_i} 2\pi(-\frac{1}{\beta})\Big\{(\frac{1}{\alpha_j})\psi^{(\frac{1}{\alpha_j}-1)}\exp(-\beta\psi^{\frac{1}{\alpha_j}}) + \psi^{\frac{1}{\alpha_j}}\exp(-\beta\psi^{\frac{1}{\alpha_j}})(-\beta\frac{1}{\alpha_j})\psi^{(\frac{1}{\alpha_j}-1)}\Big\} \\
&\qquad +\lambda_m p_{m_i} 2\pi(-\frac{1}{\beta^2})\exp(-\beta\psi^{\frac{1}{\alpha_j}})\big(-\beta\frac{1}{\alpha_j}\psi^{(\frac{1}{\alpha_j}-1)}\big)\Big\} + \lambda_m p_{m_i}\pi\frac{2}{\alpha_{\mathcal{N}}}\psi^{\frac{2}{\alpha_{\mathcal{N}}}-1} \\
&= \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} c_j\lambda_m p_{m_i} 2\pi(\frac{1}{\alpha_j})\exp(-\beta\psi^{\frac{1}{\alpha_j}})\psi^{(\frac{2}{\alpha_j}-1)} + \lambda_m p_{m_i}\pi\frac{2}{\alpha_{\mathcal{N}}}\psi^{\frac{2}{\alpha_{\mathcal{N}}}-1} \\
&\overset{(b)}{=} \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} c_j\lambda_m p_{m_i}\pi\delta_{m_j}\exp(-\beta\psi^{\frac{\delta_{m_j}}{2}})\psi^{(\delta_{m_j}-1)} + \lambda_m p_{m_i}\pi\delta_{m_{\mathcal{N}}}\psi^{\delta_{m_{\mathcal{N}}}-1}. \qquad \text{(C.3)}
\end{aligned}$$

Here, $(a)$ follows from substituting $p_{\mathcal{L}} = \exp(-\beta r_{x_i}), p_{\mathcal{N}} = 1-\exp(-\beta r_{x_i})$ and $(b)$ follows from the fact that $\delta_{m_j} = \frac{2}{\alpha_j}$. Now, the density of the process $\Omega_i = \{\frac{r^{\alpha_j}}{\mathcal{X}_{x_i}} (\triangleq \frac{||y_{i,q}||^{\alpha_j}}{\mathcal{X}_{x_i}})\} = \{||\omega_q||\}$ is acquired according to the displacement theorem. Below we show its derivation. But, first we give the joint probability of $\psi_q$ and $\mathcal{X}_{x_i}$ as

$$\mathbb{P}[\frac{\psi}{\mathcal{X}_{x_i}} \le \omega] = \mathbb{P}[\mathcal{X}_{x_i} \ge \frac{\psi}{\omega}] = 1 - F_{\mathcal{X}}(\frac{\psi}{\omega}). \qquad \text{(C.4)}$$

Due to the fact that the integral of pdf is its cdf, the joint probability is given by

$$f(\psi,\omega) = \frac{\mathrm{d}\big(1-F_{\mathcal{X}}(\frac{\psi}{\omega})\big)}{\mathrm{d}\omega} = \frac{\psi}{\omega^2} f_{\mathcal{X}}(\frac{\psi}{\omega}) = \frac{\psi}{\omega^2}\frac{\hat{m}^{\hat{m}}(\frac{\psi}{\omega})^{\hat{m}-1}\exp(-\hat{m}(\frac{\psi}{\omega}))}{\Gamma(\hat{m})}. \qquad \text{(C.5)}$$

According to the displacement theorem, we use the above joint probability to calculate the density of the process $\Omega_i$, which is given by

$$\begin{aligned}
\lambda_{\Omega_i}(\omega) &= \int_0^\infty \lambda(\psi) f(\psi,\omega)\mathrm{d}\psi \\
&= \int_0^\infty \left( \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} c_j\pi\lambda_m\exp(-\beta\psi^{\frac{\delta_{m_j}}{2}}) p_{m_i}\delta_{m_j}\psi^{(\delta_{m_j}-1)} + \lambda_m p_{m_i}\pi\delta_{m_{\mathcal{N}}}\psi^{\delta_{m_{\mathcal{N}}}-1}\right) \\
&\qquad\qquad \times \frac{\psi\hat{m}^{\hat{m}}(\frac{\psi}{\omega})^{\hat{m}-1}\exp\big(-\hat{m}(\frac{\psi}{\omega})\big)}{\omega^2\Gamma(\hat{m})}\mathrm{d}\psi \\
&= \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} c_j\frac{p_{m_i}\lambda_m\delta_{m_j}\hat{m}^{\hat{m}}\pi}{\omega^{(\hat{m}+1)}\Gamma(\hat{m})}\int_0^\infty \psi^{(\delta_{m_j}+\hat{m}-1)}\exp(-\frac{\hat{m}}{\omega}\psi - \beta\psi^{\frac{\delta_{m_j}}{2}})\mathrm{d}\psi \\
&\qquad\qquad + p_{m_i}\lambda_m\pi\delta_{m_{\mathcal{N}}}\omega^{\delta_{m_{\mathcal{N}}}-1}\frac{\Gamma(\delta_{m_{\mathcal{N}}}+\hat{m})}{\hat{m}^{\delta_{m_{\mathcal{N}}}}\Gamma(\hat{m})}. \qquad \text{(C.6)}
\end{aligned}$$

Now, according to the complementary void function, the CDF of $\omega_q$ can be given as

$$F_{\Omega_i}(\tilde{\omega}) = \mathbb{P}[\omega_q < \tilde{\omega}] = 1 - \mathbb{P}\big[\Omega_i[0,\tilde{\omega}) = 0\big]. \qquad \text{(C.7)}$$

Since the displacement theorem and mapping theorem of a PPP is still a PPP, $\mathbb{P}[\Omega_i[0,\tilde{\omega}) = 0] = \exp[-\int_0^{\tilde{\omega}}\lambda_{\Omega_i}(\omega)\mathrm{d}\omega]$. Accordingly, the CDF of $\omega_q$ can be given as

$$F_{\Omega_i}(\tilde{\omega}) = 1 - \exp\Big(-\sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} k_j p_{m_i} Z_j(\tilde{\omega}) - \hat{k} p_{m_i}\tilde{\omega}^{\delta_{m_{\mathcal{N}}}}\Big), \qquad \text{(C.8)}$$

where $Z_j(\tilde{\omega}) = \int_0^\infty\int_0^{\tilde{\omega}} \frac{\exp(-\frac{\hat{m}}{\omega}\psi)}{\omega^{(\hat{m}+1)}}\mathrm{d}\omega\psi^{(\delta_{m_j}+\hat{m}-1)}\exp(-\beta\psi^{\frac{\delta_{m_j}}{2}})\mathrm{d}\psi$ and should be a positive value due to the form of the integrand, $\hat{k} = \pi\lambda_m\frac{\Gamma(\delta_{m_{\mathcal{N}}}+\hat{m})}{\Gamma(\hat{m})\hat{m}^{\delta_{m_{\mathcal{N}}}}}$, and $k_j = c_j\frac{\pi\lambda_m\hat{m}^{\hat{m}}\delta_{m_j}}{\Gamma(\hat{m})}$. Now, according to (C.1) and (C.8), the ASP of file delivery in mmWave network can be written as

$$\mathbb{P}[\gamma^{\mathrm{SNR}}_{m_{x_i}} \ge Q_i] = F_{\Omega_i}(\frac{\mathrm{P}_m G_{x_i}}{Q_i\sigma_m^2}) = 1 - \exp\Big(-\sum_{j\in\{\mathcal{L},\mathcal{N}\}} k_j p_{m_i} Z_j(\frac{\mathrm{P}_m G_{x_i}}{Q_i\sigma_m^2}) - \hat{k} p_{m_j}(\frac{\mathrm{P}_m G_{x_i}}{Q_i\sigma_m^2})^{\delta_{m_{\mathcal{N}}}}\Big) \qquad \text{(C.9)}$$

Finally, we can generate the overall probability from (C.1) as

$$\mathcal{P}_s^{mm}(\nu_i) = \sum\nolimits_{i=1}^{L} f_i \Big[1 - \exp\Big( - \sum\nolimits_{j\in\{\mathcal{L},\mathcal{N}\}} k_j p_{m_i} Z_j \big(\frac{\eta_i G_{x_i}}{Q_i}\big) - \hat{k} p_{m_i} \big(\frac{\eta_i G_{x_i}}{Q_i}\big)^{\delta_{m_\mathcal{N}}} \Big)\Big], \tag{C.10}$$

where $k_j = c_j \frac{\pi \lambda_m \hat{m}^{\hat{m}} \delta_{m_j}}{\Gamma(\hat{m})}, \eta_i = \frac{\mathrm{P}_m}{\sigma_m^2}, p_j \in \{p_\mathcal{L}, p_\mathcal{N}\}, \delta_{m_j} = \frac{2}{\alpha_j}, \forall j \in \{\mathcal{L}, \mathcal{N}\}$.

## REFERENCES


[1] N. Golrezaei *et al.*, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 142–149, Apr. 2013.

[2] M. Akdeniz *et al.*, "Millimeter wave picocellular system evaluation for urban deployments," in *Proc. IEEE Global Commun. Conf. (Globecom) Workshops*, Atlanta, GA, Dec. 2013, pp. 105–110.

[3] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Comm.*, vol. 13, no. 9, pp. 5070–5083, Jun. 2014.

[4] S. Singh *et al.*, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE J. Select. Areas Commun.*, vol. 33, no. 1, pp. 2196–2211, Jan. 2015.

[5] G. R. MacCartney and T. S. Rappaport, "Rural macrocell path loss models for millimeter wave wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1663–1677, Jul. 2017.

[6] E. Turgut and M. C. Gursoy, "Coverage in heterogeneous downlink millimeter wave cellular networks," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4463–4477, Oct. 2017.

[7] J. G. Andrews *et al.*, "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Jan. 2017.

[8] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[9] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, San Francisco, CA, Mar. 2003, pp. 1543–1552.

[10] A. Zemlianov and G. de Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 3, pp. 657–667, Mar. 2005.

[11] Y. Pei, J. W. Modestino, and X. Wang, "On the throughput capacity of hybrid wireless networks using an l-maximum-hop routing strategy," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Orlando, FL, Oct. 2003, pp. 2173–2176.

[12] P. Li, C. Zhang, and Y. Fang, "Capacity and delay of hybrid wireless broadband access networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 117–125, Feb. 2009.

[13] S. Vuppala, S. Biswas, and T. Ratnarajah, "An analysis on secure communication in millimeter/micro-wave hybrid networks," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3507–3519, Aug. 2016.

[14] H. Elshaer *et al.*, "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.

[15] G. Zhang *et al.*, "Capacity of content-centric hybrid wireless networks," *IEEE Access*, vol. 5, pp. 1449–1459, Feb. 2017.

[16] J. Song, H. Song, and W. Choi, "Optimal caching placement of caching system with helpers," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, UK, Jun. 2015, pp. 1825–1830.

[17] 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.

[18] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless D2D networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 849–869, Feb. 2016.

[19] S. T. ul Hassan *et al.*, "Caching in wireless small cell networks: A storage-bandwidth tradeoff," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1175–1178, Jun. 2016.

[20] S. P. Shariatpanahi, H. Shah-Mansouri, and B. H. Khalaj, "Caching gain in wireless networks with fading: A multi-user diversity perspective," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Istanbul, Turkey, Apr. 2014, pp. 930–935.

[21] G. Paschos *et al.*, "Wireless caching: Technical misconceptions and business barriers," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 16–22, Aug. 2016.

[22] E. Bastug *et al.*, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[23] S. H. Chae, T. Q. S. Quek, and W. Choi, "Content placement for wireless cooperative caching helpers: A tradeoff between cooperative gain and content diversity gain," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6795–6807, Oct. 2017.

[24] M. Afshang and H. S. Dhillon, "Optimal geographic caching in finite wireless networks," in *Proc. IEEE Signal Process. Adv. Wireless Commun. (SPAWC)*, Edinburgh, UK, Jul. 2016, pp. 1–5.

[25] J. Rao *et al.*, "Optimal caching placement for D2D assisted wireless caching networks," in *Proc. IEEE Int. Conf. Commun.(ICC)*, Kuala Lumpur, Malaysia, May. 2016, pp. 1–6.

[26] Z. Chen *et al.*, "Cooperative caching and transmission design in cluster-centric small cell networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3401–3415, May. 2016.

[27] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Feb. 2015.

[28] 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.

[29] S. Biswas *et al.*, "On the performance of relay aided millimeter wave networks," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 576–588, Apr. 2016.

[30] J. Fan, Z. Xu, and G. Y. Li, "Performance analysis of MU-MIMO in downlink cellular networks," *IEEE Communications Letters*, vol. 19, no. 2, pp. 223–226, Feb. 2015.

[31] H. S. Jo *et al.*, "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.

[32] D. Maamari, N. Devroye, and D. Tuninetti, "Coverage in mmwave cellular networks with base station co-operation," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2981–2994, Apr. 2016.

[33] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.

[34] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Communi.*, vol. 12, no. 5, pp. 2484–2497, May. 2013.

[35] X. Shen *et al.*, "Disciplined convex-concave programming," in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, Las Vegas, NV, Dec. 2016, pp. 1009–1014.

[36] E. Gilbert, "Random subdivisions of space into crystals," *The Annals of mathematical statistics*, vol. 33, no. 3, pp. 958–972, Sep. 1962.