D2D-Aware Device Caching in MmWave-Cellular Networks

Nikolaos Giatsoglou, *Member, IEEE*, Konstantinos Ntontin, *Member, IEEE*, Elli Kartsakli, *Senior Member, IEEE*, Angelos Antonopoulos, *Senior Member, IEEE*, and Christos Verikoukis, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a novel policy for device caching that facilitates popular content exchange through high-rate device-to-device (D2D) millimeter-wave (mmWave) communication. The D2D aware caching (DAC) policy splits the cacheable content into two content groups and distributes it randomly to the user equipment devices (UEs), with the goal to enable D2D connections. By exploiting the high-bandwidth availability and the directionality of mmWaves, we ensure high rates for the D2D transmissions, while mitigating the co-channel interference that limits the D2D-communication potentials in the sub-6 GHz bands. Furthermore, based on a stochastic-geometry approach for the modeling of the network topology, we analytically derive the offloading gain that is achieved by the proposed policy and the distribution of the content retrieval delay considering both half- and full-duplex mode for the D2D communication. The accuracy of the proposed analytical framework is validated through Monte-Carlo simulations. In addition, for a wide range of a content popularity indicator the results show that the proposed policy achieves higher offloading and lower content-retrieval delays than existing state-of-the-art approaches.**

***Index Terms*—caching policies, Zipf popularity model, stochastic-geometry, full-duplex communication.**

I. Introduction

A. Background

OVER the last few years, the proliferation of mobile devices connected to the Internet, such as smartphones and tablets, has led to an unprecedented increase in wireless traffic that is expected to grow with an annual rate of 53% until 2020 [1]. To satisfy this growth, a goal has been set for the 5th generation (5G) of mobile networks to improve the capacity of current networks by a factor of 1000 [2]. While traditional approaches improve the area spectral efficiency of the network through, e.g., cell densification, transmission in the millimeter-wave (mmWave) band, and massive MIMO [2], studies have highlighted the repetitive pattern of user content requests [3], [4], suggesting more efficient ways to serve them.

With *proactive caching*, popular content is stored inside the network during off-peak hours (e.g., at night), so that it can be served locally during peak hours [5]. Two methods are distinguished in the literature: i) *edge caching* [6] when the content is stored at helper nodes, such as small-cell base stations (BSs), and ii) *device caching* [7] when the content is stored at the user equipment devices (UEs). While edge caching alleviates the backhaul constraint of the small-cells by reducing the transmissions from the core network, device caching offloads the BSs by reducing the cellular transmissions, which increases the rates of the active cellular UEs and reduces the dynamic energy consumption of the BSs [8]. The UEs also experience lower delays since the cached content is served instantaneously or through D2D communication from the local device caches.

The benefits of device caching in the offloading and the throughput performance have been demonstrated in [7], [9]–[13]. In [7], the spectrum efficiency of a network of D2D UEs that cache and exchange content from a content library, is shown to scale linearly with the network size, provided that their content requests are sufficiently redundant. In [9], the previous result is extended to the UE throughput, which, allowing for a small probability of outage, is shown to scale proportionally with the UE cache size, provided that the aggregate memory of the UE caches is larger than the library size. To achieve these scaling laws, the impact of the D2D interference must be reduced by optimally adjusting the D2D transmission range to the UE density. In [10], a cluster-based approach is proposed to address the D2D interference where the D2D links inside a cluster are scheduled with time division multiple access (TDMA). The results corroborate the scaling of the spectrum efficiency that was derived in [7]. In [11], a mathematical framework based on stochastic geometry is proposed to analyze the cluster-based TDMA scheme, and the trade-off between the cluster density, the local offloading from inside the cluster, and the global offloading from the whole network is demonstrated through extensive simulations. In [12], the system throughput is maximized by jointly optimizing the D2D link scheduling and the power allocation, while in [13], the offloading is maximized by a reactive caching mechanism where the UEs make caching decisions by estimating the popularity of the received content.

Although the aforementioned works show positive results for device caching, elaborate scheduling and power allocation schemes are required to mitigate the D2D interference, which limit the UE throughput and increase the system complexity. The high impact of the D2D interference is attributed to the omni-directional transmission patterns that are commonly employed in the sub-6 GHz bands. While directionality could naturally mitigate the D2D interference and alleviate the need for coordination, it requires a large number of antennas, whose size is not practical in the microwave bands. In contrast, the

N. Giatsoglou and E. Kartsakli are with IQUADRAT Informatica S.L., Spain (e-mail: ngiatsoglou@iquadrat.com, ellik@iquadrat.com)

K. Ntontin is with the Department of Informatics and Telecommunications, University of Athens (UoA), Greece (e-mail: kntontin@di.uoa.gr).

A. Antonopoulos and C. Verikoukis are with the Telecommunications Technological Centre of Catalonia (CTTC/CERCA), Spain (e-mail: aantonopoulos@cttc.es,cveri@cttc.es)

mmWave bands allows the employment of antenna arrays in hand-held UE devices due to their small wavelength. Combined with the availability of bandwidth and their prominence in future cellular communications [2], the mmWave bands are an attractive solution for D2D communication [14], [15].

The performance of the mmWave bands in wireless communication has been investigated in the literature for both outdoor and indoor environments, especially for the frequencies of 28 and 73 GHz that exhibit small atmospheric absorption [16], [17]. According to these works, the coverage probability and the average rate can be enhanced with dense mmWave deployments when highly-directional antennas are employed at both the BSs and the UEs. MmWave systems further tend to be noise-limited due to the high bandwidth and the directionality of communication [18]. Recently, several works have conducted system-level analyses of mmWave networks with stochastic geometry [19]–[21], where the positions of the BSs and the UEs are modeled according to homogeneous Poisson point processes (PPPs) [22]. This modeling has gained recognition due to its tractability [23].

B. Motivation and Contribution

Based on the above, it is seen that device caching can significantly enhance the offloading and the delay performance of the cellular network, especially when the UEs exchange cached content through D2D communication. On the other hand, the D2D interference poses a challenge in conventional microwave deployments due to omni-directional pattern of transmission. While directionality is difficult to achieve in the sub-6 GHz band for hand-held devices, it is practical in the mmWave frequencies, due to the small size of the antennas. The high availability of bandwidth and the prominence of the mmWave bands in future cellular networks have further motivated us to consider mmWave D2D communication in a device caching application. To the best of our knowledge, this combination has only been considered in [24], which adopts a cluster-based TDMA approach for the coordination of the D2D links and does not exploit the directionality of mmWaves to further increase the D2D frequency reuse.

In this context, the contributions of our work are summarized as follows:

- We propose a novel D2D-aware caching (DAC) policy for device caching that facilitates the content exchange between the paired UEs through mmWave D2D communication and exploits the directionality of the mmWave band to increase the frequency reuse among the D2D links. In addition, we consider a half-duplex (HD) and a full-duplex (FD) version of the DAC policy when simultaneous requests occur inside a D2D pair.
- We evaluate the performance of the proposed policy in terms of an offloading metric and the distribution of the content retrieval delay, based on a stochastic geometry framework for the positions of the BSs and the UEs.
- We compare our proposal with the state-of-the-art most-popular content (MPC) policy through analysis and simulation, which shows that our policy improves the offloading metric and the 90-th percentile of the delay when the availability of paired UEs is sufficiently high and the content popularity distribution is not excessively peaked.

The rest of the paper is organized as follows. In Section II, we present the proposed DAC and the state-of-the-art MPC policy. In Section III, we present the system model. In Section IV and Section V, we characterize the performance of the two policies in terms of the offloading factor and the content retrieval delay, respectively. In Section VI, we compare analytically and through simulations the performance of the caching policies. Finally, Section VII concludes the paper.

II. Background and Proposed Caching Policy

In this section, based on a widely considered model for device caching, we present the state-of-the-art MPC policy and the proposed DAC policy.

A. Device Caching Model

We assume that the UEs request content from a library of L files of equal size σ_{file} [6] and that their requests follow the Zipf distribution. According to this model, after ranking the files with decreasing popularity, the probability q_i of a UE requesting the i-th ranked file is given by

$$q_i = \frac{i^{-\xi}}{\sum_{j=1}^{L} j^{-\xi}}, \ 1 \leq i \leq L, \ \xi \geq 0, \tag{1}$$

where ξ is the popularity exponent of the Zipf distribution. This parameter characterizes the skewness of the popularity distribution and depends on the content[1] type, (e.g., webpages, video, audio, etc.) [25], [26].

In device caching, every UE retains a cache of K files, where $K << L$, so that when a cached content is requested, it is retrieved locally with negligible delay instead of a cellular transmission. This event is called a *cache hit* and its probability is called the *hit probability*, which is denoted by h and given by

$$h = \sum_{i \in C} q_i, \tag{2}$$

where C represents the cached contents of a UE, as determined by the caching policy.

B. State-of-the-Art MPC Policy

The MPC policy is a widely considered caching scheme [10], [27], [28] that stores the K most popular contents from the library of L files in every UE. This content placement maximizes the hit probability, which is given by

$$h_{mpc} = \sum_{i=1}^{K} q_i = \frac{\sum_{i=1}^{K} i^{-\xi}}{\sum_{j=1}^{L} j^{-\xi}}. \tag{3}$$

C. Proposed DAC Policy

Although the MPC policy maximizes the hit probability, it precludes content exchange among the UEs since all of them store the same files. In contrast, a policy that diversifies

[1] Please note that the terms *file* and *content* are used interchangeably in the following.

the content among the UEs enables the content exchange through D2D communication, resulting in higher offloading. Furthermore, thanks to the high D2D rate and the enhancement in the cellular rate due to the offloading, the considered policy may also improve the content retrieval delay, despite its lower hit probability compared with the MPC policy.

Based on this intuition, in the proposed DAC policy, the $2K$ most popular contents of the library of L files are partitioned into two non-overlapping groups of K files, denoted by groups A and B, and are distributed randomly to the UEs, which are characterized as UEs A and B respectively. When a UE A is close to a UE B, the network may pair them to enable content exchange through D2D communication. Denoting by h_A and h_B the hit probabilities of the two UE types, three possibilities exist when a paired UE A requests content:

- the content is retrieved through a cache hit from the local cache of UE A with probability h_A.
- the content is retrieved through a D2D transmission from the cache of UE B with probability h_B.
- the content is retrieved through a cellular transmission from the associated BS of UE A with probability $1 - h_A - h_B$.

The above cases are defined accordingly for UE B. In Proposition 1 that follows, we formally prove that the probability of content exchange for both paired UEs are maximized with the content assignment of the DAC policy.

Proposition 1: Denoting by C_A and C_B the caches of UE A and B inside a D2D pair, and by e_A and e_B their probabilities of content exchange, e_A and e_B are maximized when C_A and C_B form a non-overlapping partition of the $2K$ most popular contents, i.e., $C_A \cup C_B = \{i \in \mathbb{N} : 1 \leq i \leq 2K\}$ and $C_A \cap C_B = \emptyset$, in the sense that no other content assignment to C_A and C_B can *simultaneously* increase e_A and e_B.

Proof: see [29, Appendix A] □

When the caches of UEs A and B are non-overlapping, the hit probabilities of two paired UEs coincide with their content exchange probabilities, i.e., $e_A = h_B$ and $e_B = h_A$, hence, the DAC policy also maximizes h_A and h_B over all possible $2K$ partitions in the sense of Proposition 1[2]. The $2K$ most popular contents can be further partitioned in multiple ways, but one that equalizes h_A and h_B is chosen for fairness considerations. Although exact equalization is not possible due to the discrete nature of the Zipf distribution, the partition that minimizes the difference $|h_A - h_B|$ can be found. Considering that this difference is expected to be negligible for sufficiently high values of K, h_A and h_B can be expressed as

$$h_A \approx h_B \approx h_{dac} = \frac{1}{2}\sum_{i=1}^{2K} q_i. \tag{4}$$

Finally, since two paired UEs may want to simultaneously exchange content, with probability h_{dac}^2, we consider two cases for the DAC policy: i) an HD version, denoted by HD-DAC, where the UEs exchange contents with two sequential HD transmissions, and ii) an FD version, denoted by FD-DAC, where the UEs exchange contents simultaneously with one FD transmission. Although the FD-DAC policy increases the frequency reuse of the D2D transmissions compared with the HD-DAC policy, it also introduces self-interference (SI) at the UEs that operate in FD mode and increases the D2D co-channel interference. It therefore raises interesting questions regarding the impact of FD communication on the rate performance, especially in a mmWave system where the co-channel interference is naturally mitigated by the directionality.

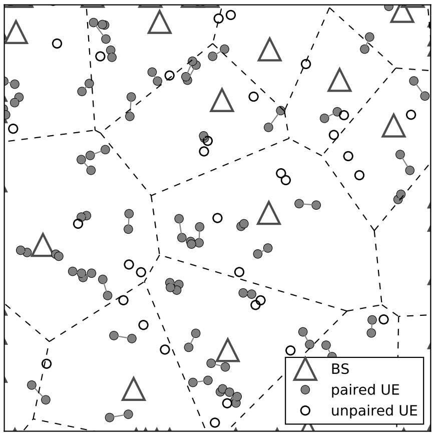


Fig. 1: A network snapshot in a rectangle of dimensions 300 m x 300 m consisting of BSs (triangles) and UEs (circles). The paired UEs are shown connected with a solid line.

III. System Model

In this section, we present the network model, the mmWave channel model, the FD operation of the UEs, and the resource allocation scheme for the cellular and the D2D transmissions.

A. Network Model

We consider a cellular network where a fraction of the UEs are paired, as shown in the snapshot of Fig. 1. We assume that the BSs are distributed on the plane according to a homogeneous PPP Φ_{bs} of intensity λ_{bs}, while the UEs are distributed according to three homogeneous PPPs: the PPP Φ_u with intensity λ_u representing the unpaired UEs, and the PPPs $\Phi_p^{(1)}$ and $\Phi_p^{(2)}$ with the same intensity λ_p representing the paired UEs. We assume that Φ_u is independent of $\Phi_p^{(1)}$ and $\Phi_p^{(2)}$, while $\Phi_p^{(1)}$ and $\Phi_p^{(2)}$ are dependent due to the correlation introduced by the D2D pairings. Specifically, for every UE of $\Phi_p^{(1)}$, a D2D peer exists in $\Phi_p^{(2)}$ that is uniformly distributed inside a disk of radius r_{d2d}^{max}, or, equivalently, at a distance r_{d2d} and angle ϕ_{d2d} that are distributed according to the probability density functions (PDFs) $f_{r_{d2d}}(r)$ and $f_{\phi_{d2d}}(\phi)$, given by

$$f_{r_{d2d}}(r) = \frac{2r}{(r_{d2d}^{max})^2},\ 0 < r < r_{d2d}^{max}, \tag{5a}$$

$$f_{\phi_{d2d}}(\phi) = \frac{1}{2\pi},\ 0 \leq \phi < 2\pi. \tag{5b}$$

We assume that the D2D pairings arise when content exchange is possible, based on the cached files of the UEs. In the DAC policy, the BSs distribute the content groups A and B

[2] Please note that h_A and h_B are still lower than h_{mpc}, since the MPC policy is not based on partitions.

independently and with probability 1/2 to their associated UEs, and a fraction δ of them, which are located within distance r_{d2d}^{max}, are paired. Defining the aggregate process of the UEs Φ_{ue} as

$$\Phi_{ue} \triangleq \Phi_u \cup \Phi_p^{(1)} \cup \Phi_p^{(2)}, \tag{6}$$

and its intensity λ_{ue} as[3]

$$\lambda_{ue} = \lambda_u + 2\lambda_p, \tag{7}$$

the ratio δ of the paired UEs is given by

$$\delta = \frac{2\lambda_p}{\lambda_{ue}} = \frac{2\lambda_p}{\lambda_u + 2\lambda_p}. \tag{8}$$

Regarding the UE association, we assume that all the UEs are associated with their closest BS[4], in which case the cells coincide with the Voronoi regions generated by Φ_{bs}. Denoting by A_{cell} the area of a typical Voronoi cell, the equivalent cell radius r_{cell} is defined as

$$r_{cell} \triangleq \sqrt{\frac{\mathbb{E}[A_{cell}]}{\pi}} = \frac{1}{\sqrt{\pi\lambda_{bs}}}, \tag{9}$$

and the association distance r of a UE to its closest BS is distributed according to the PDF $f_r(r)$, given by

$$f_r(r) = \frac{2r}{r_{cell}^2} e^{-\left(\frac{r}{r_{cell}}\right)^2} = 2\lambda_{bs}\pi r e^{-\lambda_{bs}\pi r^2}, \; r > 0. \tag{10}$$

This result follows from the void probability of the PPP [23].

B. Channel Model

Regarding the channel model, we assume that the BSs and the UEs transmit with constant power, which is denoted by P_{bs} and P_{ue} respectively, and consider mmWave transmission for both the cellular and the D2D communication through directional antennas employed at both the BSs and the UEs. The antenna gains are modeled according to the sectorized antenna model [30], which assumes constant mainlobe and sidelobe gains, given by

$$G_i(\theta) = \begin{cases} G_i^{max} & \text{if } |\theta| \leq \Delta\theta_i, \quad \text{(11a)} \\ G_i^{min} & \text{if } |\theta| > \Delta\theta_i, \quad \text{(11b)} \end{cases}$$

where $\Delta\theta$ is the antenna beamwidth, θ is the angle deviation from the antenna boresight, and $i \in \{bs, ue\}$.

Because the mmWave frequencies are subject to blockage effects, which become more pronounced as the transmission distance increases [16], the line-of-sight (LOS) state of the mmWave links is explicitly modeled. We consider the *exponential model* [16], [20], according to which a link of distance r is LOS with probability $\mathrm{P}_{los}(r)$ or non-LOS (NLOS) with probability $1 - \mathrm{P}_{los}(r)$, where $\mathrm{P}_{los}(r)$ is given by

$$\mathrm{P}_{los}(r) = e^{-\frac{r}{r_{los}}}. \tag{12}$$

The parameter r_{los} is the average LOS radius, which depends on the size and the density of the blockages [20]. We further assume that the LOS states of different links are independent and that the shadowing is incorporated into the LOS model [31]. Finally, we assume Rayleigh fast fading where the channel power gain, denoted by η, is exponentially distributed, i.e., $\eta \sim Exp(1)$.

C. FD-Operation Principle

When a UE operates in FD mode, it receives SI by its own transmission. The SI signal comprises a direct LOS component, which can be substantially mitigated with proper SI cancellation techniques, and a reflected component, which is subject to multi-path fading. Due to the lack of measurements regarding the impact of the aforementioned components in FD mmWave transceivers, we model the SI channel as Rayleigh [32], justified by the reduction of the LOS component due to the directionality [33]. Denoting by η_{si} the power gain of the SI channel including the SI cancellation scheme, and by κ_{si} its mean value, i.e, $\kappa_{si} = \mathbb{E}[\eta_{si}]$, the power of the remaining SI signal I_{si} is given by

$$I_{si} = \eta_{si} P_{ue}, \tag{13}$$

where $\eta_{si} \sim \mathrm{Exp}\left(\frac{1}{\kappa_{si}}\right)$.

D. Resource Allocation and Scheduling

We focus on the downlink of the cellular system, which is isolated from the uplink through frequency division depluxing (FDD), since the uplink performance is not relevant for the considered caching scenario. We further consider an inband overlay scheme for D2D communication [34], where a fraction χ_{d2d} of the overall downlink spectrum BW is reserved for the D2D traffic, justified by the availability of spectrum in the mmWave band. Regarding the scheduling scheme, we consider TDMA scheduling for the active cellular UEs, which is suited to mmWave communication [35], and uncoordinated D2D connunication for the D2D UEs, relying on the directionality of the mmWave transmissions for the interference mitigation.

IV. Offloading Analysis

In this section, the DAC and the MPC policies[5] are compared in terms of their offloading performance, which can be quantified by the *offloading factor* F, defined as the ratio of the average offloaded requests (i.e., requests that are not served through cellular connections) to the total content requests in the network, i.e.

$$F \triangleq \frac{\mathbb{E}[\text{offloaded requests}]}{\text{total requests}}. \tag{14}$$

The offloading factor F is derived for each policy as follows:

- In the MPC policy, a content request can be offloaded only through a cache hit, hence

$$F_{mpc} = h_{mpc}. \tag{15}$$

[3] Please note that Φ_{ue} is not a PPP due to the correlation introduced by the processes of the paired UEs, Φ_p^1 and Φ_p^2. Nevertheless, its intensity can still be defined as the average number of UEs per unit area.

[4] Different cell-association criteria, such as selecting the BS that offers the maximum received power, could have been considered, but the trends regarding the comparison of the proposed policy with state-of-the-art schemes are not expected to be affected. Hence, we consider the closest BS cell association due to its analytical tractability.

[5] Note that the same network topology, as described in Section III-A, has been assumed for both schemes to ensure a fair performance comparison.

- In the DAC policy, in addition to a cache hit, a content request of a paired UE can be offloaded through D2D communication, hence

$$F_{dac} = \delta \cdot 2h_{dac} + (1-\delta)\cdot h_{dac} = (1+\delta)h_{dac}. \quad (16)$$

Based on the above, the relative gain of the DAC over the MPC policy in terms of the offloading factor, denoted by F_{gain}, is given by

$$F_{gain} = \frac{F_{dac}}{F_{mpc}} = (1+\delta)h_{ratio}, \quad (17)$$

where h_{ratio} represents the ratio of the hit probabilities of the two policies, given by

$$h_{ratio} = \frac{h_{dac}}{h_{mpc}} = \frac{1}{2}\frac{\sum_{i=1}^{2K} i^{-\xi}}{\sum_{j=1}^{K} j^{-\xi}}. \quad (18)$$

We observe that F_{gain} depends on the fraction of the paired UEs δ, the UE cache size K and the content popularity exponent ξ, but not the library size L. The impact of K and ξ on h_{ratio} and, consequently, F_{gain} is analytically investigated in Proposition 2 that follows.

Proposition 2: The ratio of the hit probabilities of the two policies (h_{ratio}) decreases monotonically with the popularity exponent ξ and the UE cache size K. In addition, the limit of h_{ratio} with high values of K is equal to

$$\lim_{K\to\infty} h_{ratio} = \max\left(2^{-\xi}, \frac{1}{2}\right). \quad (19)$$

Proof: See [29, Appendix B]. □

Proposition 2 implies that h_{ratio} attains its minimum value for $\xi \longrightarrow \infty$, and its maximum value for $\xi = 0$, hence

$$\frac{1}{2} < h_{ratio} \leq 1 \implies \frac{1+\delta}{2} < F_{gain} \leq 1+\delta. \quad (20)$$

This result shows that for $\delta = 1$, representing the case of a fully paired network, the DAC policy always exhibits higher offloading than the MPC policy, while for $\delta = 0$, representing the case of a fully unpaired network, the converse holds. For an intermediate value of δ, the offloading comparison depends on ξ and K and is determined through (17). Finally, in Fig. 2, the convergence of h_{ratio} to its limit value for high values of K is depicted. This limit is a lower bound to h_{ratio} and serves as a useful approximation, provided that ξ is not close to 1 because, in this case, the convergence is slow.

V. Performance Analysis

In this section, the DAC and the MPC policy are characterized in terms of their rate and delay performance. The complementary CDF (CCDF) of the cellular rate is derived in Section V-A, the CCDF of the D2D rate is derived in Section V-B, and the CDFs of the content retrieval delay for both policies is derived in Section V-C.

A. Cellular Rate Analysis

Justified by the stationarity of the PPP [22], we focus on a *target UE* inserted at the origin of the network and derive the experienced cellular rate, denoted by $\mathcal{R}_{cell}$, when an uncached content is requested. The rate $\mathcal{R}_{cell}$ is determined by the cellular signal-to-interference-ratio (SINR), denoted by $SINR_{cell}$, and the load of the associated cell, denoted by $\mathcal{N}_{cell}$, through the Shannon capacity formula, modified to include the effect of the TDMA scheduling as [36]

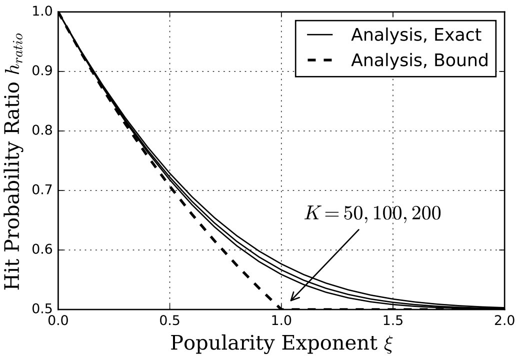


Fig. 2: The hit probability ratio h_{ratio} in terms of the UE cache size K and the popularity exponent ξ.

$$\mathcal{R}_{cell} = \frac{BW_{cell}}{\mathcal{N}_{cell}} \log\left(1 + SINR_{cell}\right) \text{ [bps].} \quad (21)$$

Based on (21), the distribution of $\mathcal{R}_{cell}$ is derived through the distribution of $SINR_{cell}$ and $\mathcal{N}_{cell}$ as

$$\mathrm{P}(\mathcal{R}_{cell} > \rho) = \mathrm{P}\left(SINR_{cell} > 2^{\frac{\rho \mathcal{N}_{cell}}{BW_{cell}}} - 1\right) =$$
$$= \sum_{n=1}^{\infty} \mathrm{P}(\mathcal{N}_{load} = n)\mathrm{P}\left(SINR_{cell} > 2^{\frac{\rho n}{BW_{cell}}} - 1 \middle| \mathcal{N}_{load} = n\right) \overset{(i)}{\approx}$$
$$\approx \sum_{n=1}^{\infty} \mathrm{P}(\mathcal{N}_{load} = n)\mathrm{P}\left(SINR_{cell} > 2^{\frac{\rho n}{BW_{cell}}} - 1\right), \quad (22)$$

where (i) follows by treating $SINR_{cell}$ and $\mathcal{N}_{cell}$ as independent random variables[6], while the distributions of $\mathcal{N}_{cell}$ and $SINR_{cell}$ are derived in the following sections.

1) Distribution of the cellular load: The distribution of $\mathcal{N}_{load}$ depends on the cell size A_{cell} and the point process of the active cellular UEs, denoted by Φ_{cell}, as follows:

- Regarding A_{cell}, we note that due to the closest BS association scheme, the cells coincide with the Voronoi regions of Φ_{bs}. Although the area distribution of a typical 2-dimensional Voronoi cell is not known, it can be accurately approximated by [37],

$$f_{A_{cell}}(a) \approx \frac{(\lambda_{bs}\kappa)^{\kappa} a^{\kappa-1} e^{-\kappa\lambda_{bs} a}}{\Gamma(\kappa)}, \ \kappa = 3.5. \quad (23)$$

The cell of the target UE, however, is stochastically larger than a randomly chosen cell, since the target UE is more probable to associate with a larger cell, and its area

[6] Please note that $SINR_{cell}$ and $\mathcal{N}_{cell}$ are dependent, since the cell load $\mathcal{N}_{cell}$ is correlated with the size of the cell, which in turns influences both the signal received from the associated BS and the interference from the neighboring BSs. Nevertheless, this dependence cannot be modeled analytically, since the relation between the SINR and the cellular size is intractable, and is not expected to have a significant impact on $\mathcal{R}_{cell}$.

TABLE I: Cellular Probabilities

	MPC	DAC
c_u	$1 - h_{mpc}$	$1 - h_{dac}$
c_p	$1 - h_{mpc}$	$1 - 2h_{dac}$

distribution can be derived from (23) as [38]:

$$f_{A_{cell}}(a) = \frac{(\lambda_{bs}\kappa)^{\kappa+1} a^{\kappa} e^{-\kappa\lambda_{bs} a}}{\Gamma(\kappa+1)} \text{ , } \kappa = 3.5. \tag{24}$$

- Regarding Φ_{cell}, it results from the independent thinning [22] of Φ_{ue}, considering the probability of a UE being cellular. This probability is denoted by c_u and c_p for the case of an unpaired and a paired UE respectively, and its values are summarized in Table I for the two considered policies. Although Φ_{cell} is not PPP due to the correlation in the positions of the paired UEs, it can be treated as a PPP with density λ_{cell}, given by

$$\lambda_{cell} = \left[(1-\delta)\cdot c_u + \delta \cdot c_p\right]\lambda_{ue}. \tag{25}$$

This approximation is justified by the small cell radius of the mmWave BSs, which is expected to be comparable to the D2D distance of the paired UEs, so that their positions inside the cell are sufficiently randomized.

Based on the above, $\mathcal{N}_{load}$ is approximated with the number of points of one PPP that fall inside the (target) Voronoi cell of another PPP, hence, it follows the gamma-Poisson mixture distribution [38], given by

$$\mathrm{P}(\mathcal{N}_{load} = n) = \frac{\Gamma(n+\kappa)}{\Gamma(\kappa+1)\Gamma(n)}\mu^{n-1}(1-\mu)^{\kappa+1} \text{ , } n \geq 1, \tag{26}$$

where

$$\mu = \frac{\lambda_{cell}}{\kappa\lambda_{bs} + \lambda_{cell}}.$$

2) Distribution of the cellular SINR: The cellular SINR is defined as

$$SINR_{cell} \triangleq \frac{S}{I+N}, \tag{27}$$

where

- S is a random variable representing the received signal power power from the associated BS, which is located at a distance r from the target UE. Assuming that the BS and UE antennas are perfectly aligned, S is given by

$$S = \left(\frac{\bar{\lambda}_c}{4\pi}\right)^2 P_{bs} G_{bs}^{max} G_{ue}^{max} \eta r^{-a}. \tag{28}$$

- I is a random variable representing the received interference from the other-cell BSs of Φ_{bs}. Assuming that the UE density is sufficiently high, all the BSs have a UE scheduled and I is given by

$$I = \sum_{x\in\Phi_{bs}} \left(\frac{\bar{\lambda}_c}{4\pi}\right)^2 P_{bs} G_x \eta_x r_x^{-a_x}, \tag{29}$$

where r_x and G_x are the length and the gain of the interfering link respectively. The latter comprises the antenna gains of the interfering BS and the target UE.

- N is the noise power at the receiver, given by

$$N = N_0 F_N BW_{cell}, \tag{30}$$

where N_0 is the noise power density, F_N is the noise figure of the receiver, and BW_{cell} is the cellular bandwidth.

Introducing the normalized quantities

$$\begin{aligned} g_x &\triangleq \frac{G_x}{max(G_x)} = \frac{G_x}{G_{bs}^{max} G_{ue}^{max}}, \\ \hat{S} &\triangleq \eta r^{-a}, \\ \hat{I} &\triangleq \sum_{x\in\Phi_{bs}} g_x \eta_x r_x^{-a_x}, \\ \hat{N} &\triangleq \left(\frac{4\pi}{\bar{\lambda}_c}\right)^2 \frac{N_0 F_N BW_{cell}}{P_{bs} G_{bs}^{max} G_{ue}^{max}}, \end{aligned} \tag{31}$$

and applying (28), (29), and (30) to (27), the expression for $SINR_{cell}$ is simplified to

$$\begin{aligned} SINR_{cell} &= \frac{\hat{S}}{\hat{I}+\hat{N}} = \\ &= \frac{\eta r^{-a}}{\sum_{x\in\Phi_{bs}} g_x \eta_x r_x^{-a_x} + \left(\frac{4\pi}{\bar{\lambda}_c}\right)^2 \frac{N_0 F_N BW_{cell}}{P_{bs} G_{bs}^{max} G_{ue}^{max}}}. \end{aligned} \tag{32}$$

The CCDF of $SINR_{cell}$ is subsequently derived as

$$\begin{aligned} \mathrm{P}\left(SINR_{cell} > T\right) &= \mathbb{E}_{r,a,\hat{I}}\left[\mathrm{P}\left(\eta > (\hat{I}+\hat{N})Tr^a\right)\right] \overset{(i)}{=} \\ &= \mathbb{E}_{r,a,\hat{I}}\left[e^{-(\hat{I}+\hat{N})Tr^a}\right] \overset{(ii)}{=} \mathbb{E}_{r,a}\left[\mathcal{L}_{\hat{I}}(Tr^a)e^{-\hat{N}Tr^a}\right], \end{aligned} \tag{33}$$

where (i) follows from the CCDF of the exponential random variable, and (ii) from the Laplace transform of $\hat{I}$. The impact of the interference and the noise to $SINR_{cell}$ is elegantly separated in (33). In the following, considering that the impact of the interference is reduced by the directionality of the mmWave transmissions, and that the impact of noise is increased due to the large bandwidth of the mmWave band, we assume that the system operates in the *noise-limited* regime, which means that $SINR_{cell}$ can be approximated by the cellular signal-to-noise-ratio (SNR), denoted by SNR_{cell}, as

$$\begin{aligned} \mathrm{P}\left(SINR_{cell} > T\right) &\approx \mathrm{P}\left(SNR_{cell} > T\right) = \mathbb{E}_{r,a}\left[e^{-\hat{N}Tr^a}\right] = \\ &= \int_0^{\infty}\left(e^{-\frac{r}{r_{los}}} e^{-\hat{N}Tr^{a_{los}}} + (1 - e^{-\frac{r}{r_{los}}})e^{-\hat{N}Tr^{a_{nlos}}}\right) f_r(r)dr, \end{aligned} \tag{34}$$

where $f_r(r)$ is given by (10). Although the integral in (34) cannot be solved in closed form, we present a tight approximation in Proposition 3 that follows.

Proposition 3: The CCDF of the cellular SINR can be accurately approximated by

$$\mathrm{P}(SINR_{cell} > T) \approx J_1(T, a_{nlos}) + J_2(T, a_{nlos}) - J_2(T, a_{los}), \tag{35}$$

where

$$J_1(T,a) = \frac{2}{a r_{cell}^2}\left(\frac{\gamma\left(\frac{2}{a}, \hat{N}Tr_1^a\right)}{(\hat{N}T)^{\frac{2}{a}}} - \frac{\gamma\left(\frac{3}{a}, \hat{N}Tr_1^a\right)}{r_1(\hat{N}T)^{\frac{3}{a}}}\right), \tag{36}$$

$$J_2(T,a) =$$
$$= \frac{2}{a r_{cell}^2}\left(\frac{\gamma\left(\frac{2}{a}, \hat{N}Tr_2^a\right)}{(\hat{N}T)^{\frac{2}{a}}} - 2\frac{\gamma\left(\frac{3}{a}, \hat{N}Tr_2^a\right)}{r_2(\hat{N}T)^{\frac{3}{a}}} + \frac{\gamma\left(\frac{4}{a}, \hat{N}Tr_2^a\right)}{r_2^2(\hat{N}T)^{\frac{4}{a}}}\right), \tag{37}$$

with

$$r_1 = \sqrt{3}r_{cell},$$
$$r_2 = \frac{\sqrt{6}r_{cell}}{r_{los}}\sqrt{1 - \sqrt{\pi}\frac{r_{cell}}{2r_{los}}e^{\left(\frac{r_{cell}}{2r_{los}}\right)^2}\mathrm{erfc}\left(\frac{r_{cell}}{2r_{los}}\right)}.$$

Proof: See [29, Appendix C]. □

B. D2D Rate Analysis

Similar to the cellular case, we focus on a paired target UE at the origin and derive the experienced D2D rate when a content is requested from the D2D peer, which is denoted by $\mathcal{R}_{d2d}$. The following analysis applies only to the DAC policy, which is distinguished for the HD-DAC and the FD-DAC policy in the following sections.

1) Distribution of the D2D rate for the HD-DAC policy: The D2D rate for the HD-DAC policy, denoted by $\mathcal{R}_{d2d}^{hd}$, is determined by the D2D SINR, denoted by $SINR_{d2d}^{hd}$, through the Shannon capacity formula as

$$\mathcal{R}_{d2d}^{hd} = \psi BW_{d2d}\log\left(1 + SINR_{d2d}^{hd}\right) \text{ [bps]}, \tag{38}$$

where ψ denotes the HD factor, equal to $1/2$ when both paired UEs want to transmit. Subsequently, the CCDF of $\mathcal{R}_{d2d}^{hd}$ is determined by the CCDF of $SINR_{d2d}^{hd}$ as

$$\mathrm{P}\left(\mathcal{R}_{d2d}^{hd} > \rho\right) = \mathrm{P}\left(SINR_{d2d}^{hd} > 2^{\frac{\rho}{\psi BW_{d2d}}} - 1\right). \tag{39}$$

Regarding $SINR_{d2d}^{hd}$, it is defined as:

$$SINR_{d2d}^{hd} \triangleq \frac{S}{I+N}, \tag{40}$$

where

- S is a random variable representing the received signal power from the D2D peer, located at a distance r_{d2d} from the target UE. Assuming that the antennas of the two UEs are perfectly aligned, S is given by
$$S = \left(\frac{\bar{\lambda}_c}{4\pi}\right)^2 P_{ue}(G_{ue}^{max})^2\eta r_{d2d}^{-a}. \tag{41}$$
- I is a random variable representing the received interference from all transmitting D2D UEs. Denoting by Φ_{d2d}^{hd} the point process of the D2D interferers in the HD-DAC policy, I is given by
$$I = \sum_{x\in\Phi_{d2d}^{hd}}\left(\frac{\bar{\lambda}_c}{4\pi}\right)^2 P_{ue}G_x\eta_x r_x^{-a_x}, \tag{42}$$
where r_x and G_x are the length and the gain of the interfering link, where the latter comprises the antenna gains of the interfering UE and the target UE. Since, in HD-DAC, at most one UE from every D2D pair can transmit, the intensity of Φ_{d2d}^{hd} is given by
$$\lambda_{d2d}^{hd} = \left(1 - (1-h_{dac})^2\right)\lambda_p = \frac{\delta}{2}h_{dac}\left(2 - h_{dac}\right)\lambda_{ue}. \tag{43}$$
- N is the noise power at the receiver, which depends on the cellular bandwidth BW_{d2d} and is given by
$$N = N_0 F_N BW_{d2d}. \tag{44}$$

Introducing the normalized quantities

$$g_x \triangleq \frac{G_x}{\max(G_x)} = \frac{G_x}{(G_{ue}^{max})^2},$$
$$\hat{S} \triangleq \eta r_{d2d}^{-a},$$
$$\hat{I} \triangleq \sum_{x\in\Phi_{bs}} g_x\eta_x r_x^{-a_x},$$
$$\hat{N} \triangleq \left(\frac{4\pi}{\bar{\lambda}_c}\right)^2\frac{N_0F_NBW_{d2d}}{P_{ue}(G_{ue}^{max})^2}, \tag{45}$$

and applying (41), (42), and (44) to (40), the expression for $SINR_{d2d}^{hd}$ can be simplified to

$$SINR_{d2d}^{hd} = \frac{\hat{S}}{\hat{I}+\hat{N}} = \frac{\eta r_{d2d}^{-a}}{\sum_{x\in\Phi_{d2d}^{hd}} g_x\eta_x r_x^{-a_x} + \left(\frac{4\pi}{\bar{\lambda}_c}\right)^2\frac{N_0F_NBW_{d2d}}{P_{ue}(G_{ue}^{max})^2}}. \tag{46}$$

Subsequently, the CCDF of $SINR_{d2d}^{hd}$ is derived similarly to (33) as

$$\mathrm{P}(SINR_{d2d}^{hd} > T) = \mathbb{E}_{r_{d2d},a}\left[\mathcal{L}_{\hat{I}}^{hd}(Tr_{d2d}^a)e^{-\hat{N}Tr_{d2d}^a}\right] =$$
$$= \int_0^{r_{d2d}^{max}} e^{-\frac{r}{r_{los}}}\mathcal{L}_{\hat{I}}^{hd}(Tr^{a_{los}})e^{-\hat{N}Tr^{a_{los}}}f_{r_{d2d}}(r)dr+ \tag{47}$$
$$+ \int_0^{r_{d2d}^{max}}\left(1 - e^{-\frac{r}{r_{los}}}\right)\mathcal{L}_{\hat{I}}^{hd}(Tr^{a_{nlos}})e^{-\hat{N}Tr^{a_{nlos}}}f_{r_{d2d}}(r)dr, \tag{48}$$

where $\mathcal{L}_{\hat{I}}^{hd}(.)$ is the Laplace transform of the D2D interference in the HD-DAC policy and $f_{r_{d2d}}(r)$ is given by (5). In contrast to the cellular case, the contribution of the interference in $SINR_{d2d}^{hd}$ is not negligible, even with directionality, due to the smaller bandwidth that is expected to be reserved for D2D communication, thus, $\mathcal{L}_{\hat{I}}^{hd}(.)$ is evaluated according to Proposition 4 that follows.

Proposition 4: The Laplace transform of the D2D interference in the HD-DAC policy $\mathcal{L}_{\hat{I}}^{hd}(.)$ is given by

$$\mathcal{L}_{\hat{I}}^{hd}(s) \approx e^{-\pi\delta\lambda_{ue}h_{dac}(2-h_{dac})\mathbb{E}_g[J_3(s,a_N)+J_4(s,a_L;k)-J_4(s,a_N;k)]}, \tag{49}$$

where

$$J_3(s,a) = \frac{1}{2}\Gamma\left(1 - \frac{2}{a_{nlos}}\right)\Gamma\left(1 + \frac{2}{a_{nlos}}\right)g^{\frac{2}{a_{nlos}}}s^{\frac{2}{a_{nlos}}},$$
$$J_4(s,a;k) = \sum_{l=0}^{k}\binom{k}{l}(-1)^l\frac{r_0^{a+2}{}_2F_1\left(1, 1+\frac{l+2}{a}; 2+\frac{l+2}{a}; -\frac{r_0^a}{gs}\right)}{(l+a+2)gs},$$
$$r_0 = \sqrt{(k+1)(k+2)}r_{los}, \tag{50}$$

k denotes the order of the approximation, and the averaging is taken over the discrete random variable g with distribution

$$g = \begin{cases} 1 & \text{with prob } \frac{\Delta\theta_{ue}^2}{4\pi^2} & \text{(51a)} \\ \dfrac{G_{ue}^{min}}{G_{ue}^{max}} & \text{with prob } 2\frac{\Delta\theta_{ue}(2\pi-\Delta\theta_{ue})}{4\pi^2} & \text{(51b)} \\ \left(\dfrac{G_{ue}^{min}}{G_{ue}^{max}}\right)^2 & \text{with prob } \frac{(2\pi-\Delta\theta_{ue})^2}{4\pi^2} & \text{(51c)} \end{cases}$$

Proof: See [29, Appendix D]. □

As $k \to \infty$, more terms are added in the summation and the approximation becomes exact. Combining (48) and (49) into (39), yields the CCDF of $\mathcal{R}_{d2d}^{hd}$ where the final integration over r_{d2d} can be evaluated numerically.

2) Distribution of the D2D rate for the FD-DAC policy: As in the case of the HD-DAC policy, the D2D rate for the FD-DAC policy, denoted by $\mathcal{R}_{d2d}^{fd}$, is determined by the D2D SINR, denoted by $SINR_{d2d}^{fd}$, through the Shannon capacity formula as

$$\mathcal{R}_{d2d}^{fd} = BW_{d2d} \log\left(1 + SINR_{d2d}^{fd}\right) \text{ [bps].} \tag{52}$$

Subsequently, the CCDF of $\mathcal{R}_{d2d}^{fd}$ is derived from the CCDF of $SINR_{d2d}^{fd}$ as

$$\mathrm{P}\left(\mathcal{R}_{d2d}^{fd} > \rho\right) = \mathrm{P}\left(SINR_{d2d}^{fd} > 2^{\frac{\rho}{BW_{d2d}}} - 1\right). \tag{53}$$

Focusing on a paired target UE at the origin, $SINR_{d2d}^{fd}$ can be expressed as

$$SINR_{d2d}^{fd} \triangleq \frac{S}{I + I_{si} + N}, \tag{54}$$

where

- S is a random variable for the received signal power from the D2D peer, given by (41).
- I_{si} is a random variable for the SI power when the target UE operates in FD mode, given by (13).
- I is the received interference power from all transmitting D2D UEs, expressed as

$$I = \sum_{x_1 \in \Phi_p^{(1)}} \left(\frac{\bar{\lambda}_c}{4\pi}\right)^2 P_{ue}\psi_{x_1} g_{x_1} \eta_{x_1} r_{x_1}^{-a_{x_1}} + \\ + \sum_{x_2 \in \Phi_p^{(2)}} \left(\frac{\bar{\lambda}_c}{4\pi}\right)^2 P_{ue}\psi_{x_2} g_{x_2} \eta_{x_2} r_{x_2}^{-a_{x_2}}, \tag{55}$$

  where $\Phi_p^{(1)}$ and $\Phi_p^{(2)}$ are the point processes of the paired UEs, and ψ_x is an indicator variable for the event that the UE at the position x transmits.
- N is the noise power at the receiver, given by (44).

Defining g, $\hat{S}$ and $\hat{N}$ as in (45) and introducing

$$\hat{I} = \sum_{x \in \Phi_p^{(1)}} \psi_x g_x \eta_x r_x^{-a_x} + \sum_{y \in \Phi_p^{(2)}} \psi_y g_y \eta_y r_y^{-a_y}, \\ \hat{I_{si}} = \left(\frac{4\pi}{\bar{\lambda}_c}\right)^2 \frac{\eta_{si}}{(G_{ue}^{max})^2}, \tag{56}$$

the CCDF of $SINR_{d2d}^{fd}$ is derived as

$$\mathrm{P}\left(SINR_{d2d}^{fd} > T\right) = \mathbb{E}_{r_{d2d},a}\left[\mathcal{L}_{\hat{I}}^{fd}(Tr_{d2d}^{a})\mathcal{L}_{\hat{I_{si}}}(Tr_{d2d}^{a})e^{-\hat{N}Tr_{d2d}^{a}}\right], \tag{57}$$

where $\mathcal{L}_{\hat{I}}^{fd}(.)$ and $\mathcal{L}_{\hat{I_{si}}}(.)$ are the Laplace transforms of the interference in the FD-DAC policy and the SI respectively. Recalling that $\eta_{si} \sim \mathrm{Exp}\left(\frac{1}{\kappa_{si}}\right)$, $\mathcal{L}_{\hat{I_{si}}}(s)$ is derived using the Laplace transform of the exponential random variable as

$$\mathcal{L}_{\hat{I_{si}}}(s) = \mathbb{E}\left[e^{-\left(\frac{4\pi}{\bar{\lambda}_c G_{ue}^{max}}\right)^2 \eta_{si} s}\right] = \frac{1}{1 + \left(\frac{4\pi}{\bar{\lambda}_c G_{ue}^{max}}\right)^2 \frac{s}{\kappa_{si}}}, \tag{58}$$

while $\mathcal{L}_{\hat{I}}^{fd}(s)$ is derived in Proposition 5 that follows.

Proposition 5: The Laplace transform of the D2D interference in the FD-DAC policy $\mathcal{L}_{\hat{I}}^{fd}$ can be bounded as

$$\mathcal{L}_{\hat{I}}^{fd}(s) \geq e^{-\pi\delta\lambda_{ue}h_{dac}2(J_3(s,a_N)+J_4(s,a_L;k)-J_4(s,a_N;k))}, \tag{59}$$

$$\mathcal{L}_{\hat{I}}^{fd}(s) \leq e^{-\pi\delta\lambda_{ue}h_{dac}(J_3(2s,a_N)+J_4(2s,a_L;k)-J_4(2s,a_N;k))}, \tag{60}$$

where $J_3(s,a)$ and $J_4(s,a;k)$ are given by (50).

Proof: See [29, Appendix E]. □

Combining (58) and (60) into (57) and applying the result to (53) yields two bounds for the CCDF of $\mathcal{R}_{d2d}^{fd}$.

C. Delay Analysis

In this section, we characterize the delay performance of the MPC and the DAC policies through the *content retrieval delay*, denoted by D and defined as the delay experienced by a UE when retrieving the requested content from any available source. In the case of a cache hit, D is zero, while in the cellular and the D2D case it coincides with the transmission delay of the content to the UE[7]. The CDFs of D for the MPC and the DAC policies are derived as follows:

- For the MPC policy, the requested content is retrieved from the local cache with probability h_{mpc}, or from the BS with probability $1 - h_{mpc}$, hence

$$\mathrm{P}(D < d) = h_{mpc} + (1 - h_{mpc})\mathrm{P}\left(\mathcal{R}_{cell} > \frac{x}{d}\right), \tag{61}$$

  where the CCDF of $\mathcal{R}_{cell}$ is given by (22).
- For the DAC policy, the case of the paired and the unpaired UE must be differentiated, since the unpaired UE lacks the option for D2D communication. For a paired UE, the requested content is retrieved from the local cache with probability h_{dac}, from the D2D peer with probability h_{dac}, or from the BS with probability $1 - 2h_{dac}$, while, for an unpaired UE, the requested content is retrieved from the local cache with probability h_{dac}, and from the BS with probability $1 - h_{dac}$, yielding

$$\mathrm{P}(D < d) = h_{dac} + \delta h_{dac}\mathrm{P}\left(\mathcal{R}_{cell} > \frac{\sigma_{file}}{d}\right) + \\ + (1 - h_{dac} - \delta h_{dac})\mathrm{P}\left(\mathcal{R}_{d2d} > \frac{\sigma_{file}}{d}\right), \tag{62}$$

[7] Additional delays caused by the retrieval of the content through the core network are out of the scope of this work.

TABLE II: SIMULATION PARAMETERS

λ_{bs}	127 BSs/km^2	N_0	-174 dBm/Hz
λ_{ue}	1270 UEs/km^2	F_N	10 dB
δ	0.5, 0.75, 1.0	$\Delta\theta_{ue}$	30^o
r_{d2d}^{max}	15 m	$\Delta\theta_{bs}$	10^o
f_c	28 GHz	G_{bs}^{max},	18 dB
BW	2 GHz	G_{bs}^{min}	-2 dB
χ_{d2d}	20%	G_{ue}^{max}	9 dB
r_{los}	30 m	G_{ue}^{min}	-9 dB
a_{los}	2	σ_{file}	100 MBs
a_{nlos}	3	L	1000
P_{bs}	30 dBm	K	50, 100, 200
P_{ue}	23 dBm	ξ	variable
κ_{si}	80 dB		

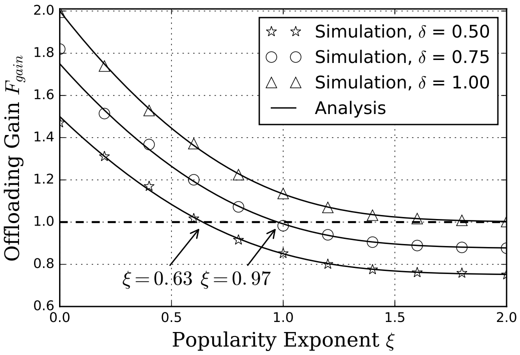


Fig. 3: The offloading gain of the DAC policy over the MPC policy F_{gain} in terms of the content popularity exponent ξ.

where the CCDF of $\mathcal{R}_{cell}$ is given by (22) and the CCDF of $\mathcal{R}_{d2d}$ is given by (39) for the HD-DAC policy and by (53) for the FD-DAC policy.

VI. Results

In this section, we compare the DAC and the MPC policy in terms of the offloading factor and the 90-th percentile of the content retrieval delay both analytically and through Monte-Carlo simulations. Towards this goal, we present the simulation parameters in Section VI-A, the results for the offloading in Section VI-B, and the results for the content retrieval delay in Section VI-C.

A. Simulation Setup

For the simulation setup of the DAC and the MPC policy, we consider a mmWave system operating at the carrier frequency of 28 GHz, which is chosen due to its favorable propagation characteristics [39] and its approval for 5G deployment by the FCC [40]. Regarding the network topology, we consider a high BS density λ_{bs} corresponding to an average cell radius r_{cell} of 50 m, which is consistent with the trends in the densification of future cellular networks and the average LOS radius r_{los} of the mmWave frequencies in urban environments [18]. The latter is chosen to be 30 m, based on the layout for the Chicago and the Manhattan area [18]. Regarding the antenna model of the BSs and the UEs, the gains and the beamwidths are chosen according to typical values of the literature [41], [42], considering that the directionality of the UE antennas will be lower due to the smaller number of antennas that can be installed in the UE devices. Regarding the caching model, we consider a library of 1000 files of size 100 MBs and three cases for the UE cache size: i) $K = 50$, ii) $K = 100$, and iii) $K = 100$, corresponding to the 5%, 10%, and 20% percentages of the library size respectively. The rest of the simulation parameters are summarized in Table II.

B. Offloading Comparison

As shown analytically in Section IV, the offloading gain of the DAC policy over the MPC policy F_{gain} increases monotonically with the UE pairing probability δ, and decreases monotonically with the UE cache size K and the content popularity ξ, while it is not affected by the library size L. In this section, we validate the impact of δ, K, and ξ on F_{gain} by means of simulations.

In Fig. 3, we plot F_{gain} in terms of ξ for $K = 100$ and for $\delta = 0.5, 0.75, 1$, corresponding to three different percentages of paired UEs inside the network. We observe that the simulation results validate the monotonic increase and decrease of F_{gain} with δ and ξ respectively. The former is attributed to the higher availability of D2D pairs, which improves the opportunities for offloading in the DAC policy and does not affect the MPC policy, while the latter is attributed to the increasing gap in the hit probabilities of the two policies, as illustrated with the decrease of h_{ratio} with ξ in Fig. 2 of Section IV. Based on the above, we observe that the maximum offloading gain of the DAC over the MPC policy is equal to 2 and it is achieved when $\delta = 1$ and $\xi = 0$, which corresponds to the case of a fully paired network and uniform content popularity respectively. For $\delta = 1$, we further observe that the DAC policy outperforms the MPC policy regardless of the value of ξ, while for lower values of δ, the DAC policy is superior only when $\xi < 0.63$ for $\delta = 0.75$, and when $\xi < 0.97$ for $\delta = 0.5$. Based on this observation, we can generalize that for a network with $\delta < 1$ the DAC policy offers higher offloading than the MPC policy for ξ up to a threshold value, which decreases with δ.

In Fig. 4, we plot the minimum δ that is required for the DAC policy to outperform the MPC policy, in terms of ξ and for $K = 50,\ 100,\ 200$. We can observe that the requirements for δ become more stringent with increasing ξ and K, which widen the gap between the hit probability of the two policies, but the impact of K is weaker than the impact of ξ, which is attributed to the low sensitivity of h_{ratio} with K. This behavior can be explained with the bound of h_{ratio} in (19), which represents the limit of h_{ratio} when $K \to \infty$. The minimum delta for this case is also depicted in Fig. 4 and the other curves converge to it for $K \to \infty$. When $\xi < 0.5$ or $\xi > 1.5$, the gap between the curves for finite K are close to the bound, because h_{ratio} converges quickly to its limit value. In contrast, when $0.5 < \xi < 1.5$, the gap between the curves and the bound is wider, because h_{ratio} converges slowly to its limit value. Due to the slow convergence, for practical values of K, similar

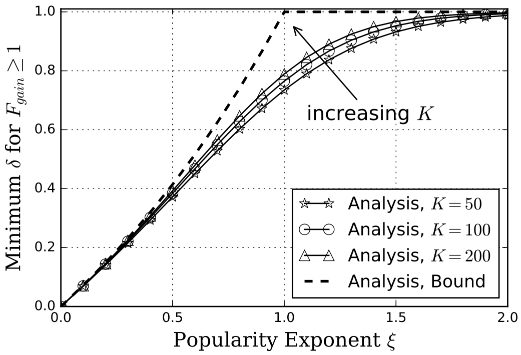


Fig. 4: The minimum fraction of pairs (δ) required for the DAC policy to achieve higher offloading than the MPC policy in terms of the content popularity exponent ξ.

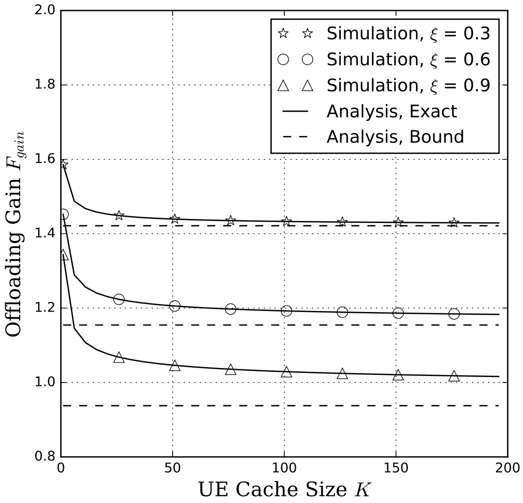


Fig. 5: The offloading gain of the DAC policy over the MPC policy F_{gain} in terms of the UE cache size K.

to ones considered in this work, h_{ratio} is insensitive to K.

In Fig. 5, we plot F_{gain} in terms of K for δ = 0.75 and ξ = 0.3, 0.6, 0.9. We observe that, as K increases, F_{gain} decreases fast at low values of K and, afterwards, tends slowly to its limit value, which is calculated by applying (19) to (17). For ξ = 0.9, the gap between the curve and the limit is high because of the slow convergence of (19), validating that F_{gain} is insensitive to K, provided that K is sufficiently high. In contrast, lower values of K favor the DAC policy over the MPC policy.

C. Delay Comparison

In this section, we validate the analytical expressions of Section V and compare the two caching policies in terms of the 90-th percentile of the content retrieval delay.

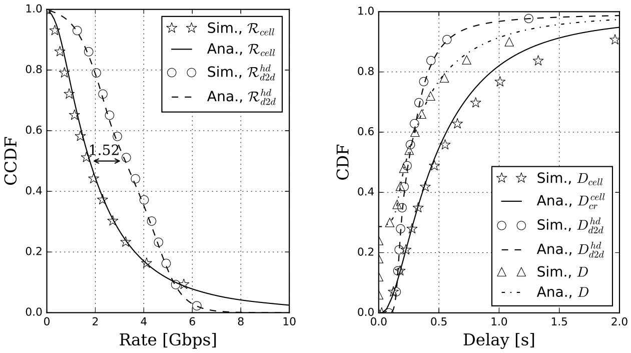
1) Performance of the HD-DAC policy: In Fig. 6a, we illustrate for the HD-DAC policy the CCDFs of the cellular rate $\mathcal{R}_{cell}$ and the D2D rate $\mathcal{R}^{hd}_{d2d}$, derived through analysis and simulations, for δ = 1, K = 200, and ξ = 0.4. We observe that $\mathcal{R}^{hd}_{d2d}$ is stochastically larger than $\mathcal{R}_{cell}$ for rates below 5 Gbps, yielding an improvement of 1.52 Gbps in the 50-th percentile, which means that the D2D UEs experience a rate that is higher than the cellular rate by at least 1 Gbps for the 50% of the time. This improvement creates strong incentives for the UEs to cooperate and is attributed to the small D2D distance between the D2D UEs and the reduction of $\mathcal{R}_{cell}$ due to the TDMA scheduling. In contrast, the cellular UEs are more probable to experience rates above 5 Gbps, owing to the high difference between the cellular and the D2D bandwidth. Specifically, it is possible for a cellular UE to associate with a BS with low or even zero load and fully exploit the cellular spectrum, while a D2D UE is always limited by the 20% fraction of bandwidth that is reserved for D2D communication.

In Fig. 6b, we illustrate for the HD-DAC policy the CDFs of the cellular delay D_{cell}, the D2D delay D^{hd}_{d2d}, and the total delay D that is experienced by a UE without conditioning on its content request. We observe that D^{hd}_{d2d} is significantly lower than D_{cell}, which is consistent with Fig. 6a, while the curve of D is initiated at the value 0.286 due to the zero delay of cache-hits. We further observe that the simulations for D_{cell} do not match the theoretical curve as tightly as in the case of $\mathcal{R}_{cell}$, which is attributed to the reciprocal relation between the rate and the delay that magnifies the approximation error for the delay. Nevertheless, the match is improved in the case of the total delay due to the contribution of the D2D delay, which is approximated more accurately.

2) Performance of the FD-DAC policy: In Fig. 7, we illustrate for the FD-DAC policy the rate and the delay distribution. As seen in Fig. 7a, the analytical bounds for the CCDF of $\mathcal{R}^{fd}_{d2d}$ are both very close to the simulation curve, hence, only one analytical bound for the D2D delay is chosen in Fig. 7b for clarity. Compared with the HD-DAC policy, the FD-DAC policy yields a minor improvement in the 50-th percentile of $\mathcal{R}^{fd}_{d2d}$, which is higher than the percentile of $\mathcal{R}_{cell}$ by 1.62 Gbps, that is attributed to the absence of the HD factor that decreases $\mathcal{R}^{hd}_{d2d}$ by half. Nevertheless, the probability of bidirectional content exchange, equal to 0.08 for the considered parameters, is small to significantly influence the results. The same observation holds for the CDFs of the content retrieval delay.

Motivated by the previous observation, in Fig. 8, we illustrate for the FD-DAC policy the rate and the content retrieval delay for ξ = 1.0, in which case h_{dac} = 0.44, resulting in a non-negligible probability for bidirectional content exchange. As seen in Fig. 8a, $\mathcal{R}^{fd}_{d2d}$ is reduced due to the higher D2D interference, while $\mathcal{R}_{cell}$ is significantly improved due to the higher offloading. Consequently, $\mathcal{R}^{fd}_{d2d}$ is higher than $\mathcal{R}_{cell}$, and the total delay is determined by the cache hits and the curve of the cellular delay, as seen in Fig. 8b. Since the FD-DAC and the HD-DAC policy are differentiated when h_{dac} is high, in which case the performance is not influenced by the D2D communication, only the HD-DAC policy is considered in the delay comparison with the MPC policy.

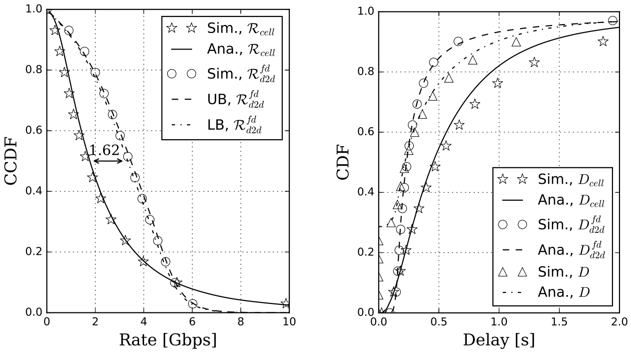
3) Delay Comparison between the MPC and the HD-DAC policy: The MPC policy maximizes the probability of zero delay through cache hits, but the HD-DAC policy may still offer lower delays due to the improvement in the transmission



(a) Rate

(b) Delay

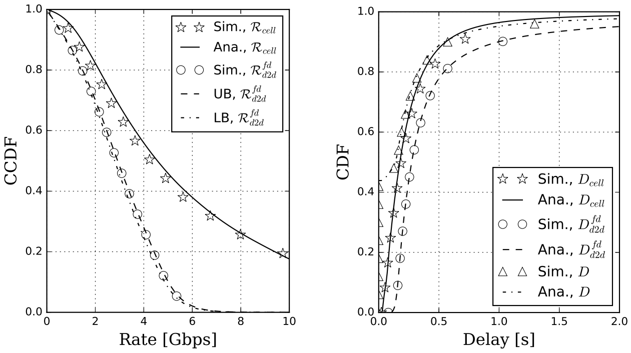
Fig. 6: Rate and delay performance of the HD-DAC policy for $K = 200$ and $\xi = 0.4$ (Ana. stands for Analysis and Sim. for Simulation).



(a) Rate

(b) Delay

Fig. 7: Rate and delay performance of the FD-DAC policy for $K = 200$ and $\xi = 0.4$ (Ana. stands for Analysis, Sim. for Simulation, UB for Upper Bound, and LB for Lower Bound).



(a) Rate

(b) Delay

Fig. 8: Rate and delay performance of the FD-DAC policy for $K = 200$ and $\xi = 1.0$ (Ana. stands for Analysis, Sim. for Simulation, UB for Upper Bound, and LB for Lower Bound).

rates. Based on this observation, the two policies are compared in terms of the 90-th percentile of the content retrieval delay, which is an important QoS metric, representing the maximum delay that is experienced by the target UE for 90% of the time.

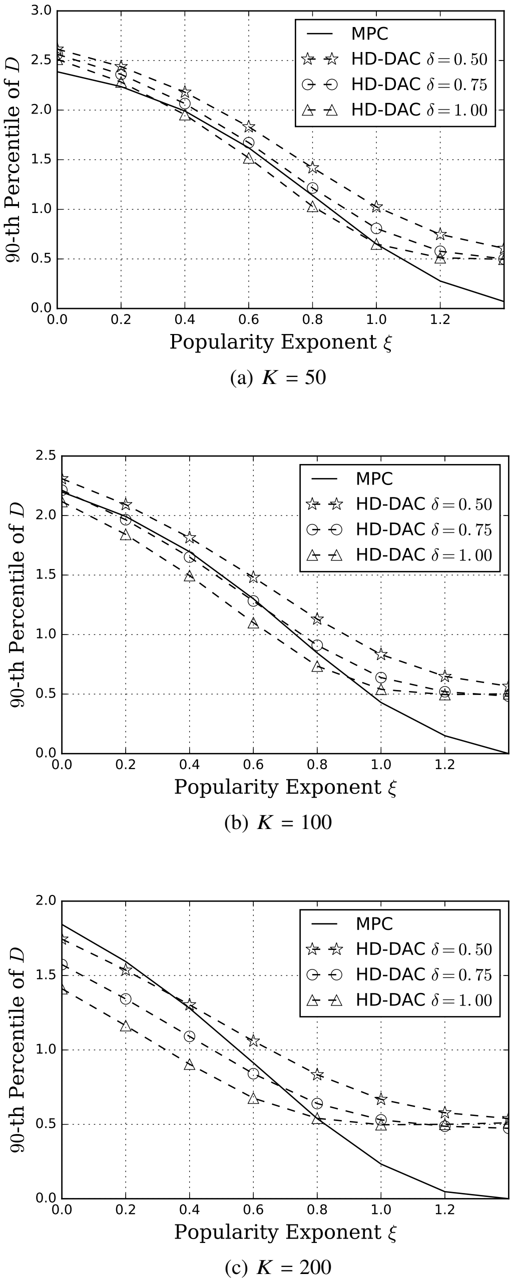
In Fig. 9, we plot the delay percentiles for the HD-DAC and the MPC policy as a function of the popularity exponent ξ for the cases: a) $K = 50$, b) $K = 100$, and c) $K = 100$, which correspond to the 5%, 10%, and 20% percentage of the library size. As a general observation, the 90-th percentile of delay for both policies decreases with higher values of K, since both the hit probability and, in the case of the HD-DAC policy, the probability of D2D content exchange, are higher. The delay percentile of the HD-DAC policy also decreases with δ, since the opportunities for D2D communication are improved with a larger number of D2D pairs, while the MPC policy is not affected. In Fig. 9a, the performance is comparable between the HD-DAC policy with $\delta = 1.0$ and the MPC policy for $\xi < 1.0$. In Fig. 9b, the performance is comparable between the HD-DAC policy with $\delta = 0.75$ and the MPC policy for $\xi < 1.0$. In Fig. 9c, the performance is comparable between the HD-DAC policy with $\delta = 1.0$ and the MPC policy for $\xi < 0.4$. Based on these these observations, we conclude that, for low values of ξ, the HD-DAC policy is favored by larger UE caches and requires fewer D2D pairings to outperform the MPC policy, while for high values of ξ, the MPC policy is favored by larger UE caches due to the wide gap in the hit probabilities of the two policies, which explains the superior performance of the MPC policy in these cases.

VII. Conclusion

Motivated by the unprecedented growth in the mobile traffic and the attractive offloading potentials of device caching, in this work, we have proposed a novel device caching policy that exploits the emerging technologies of D2D and mmWave communication to enhance the traffic offloading and the content retrieval delay of the UEs. In particular, the proposed DAC policy divides the cacheable content into two content groups and distributes it evenly at the UEs to promote content exchange between them. Based on a stochastic-geometry model, we have performed a system-level analysis and analytically derived the offloading gain and the distribution of the content retrieval delay. By comparing the proposed policy with the state-of-the-art MPC policy, which stores the most popular content to all UEs, we have showed that our approach improves the offloading and the delay performance when the availability of pairs in the system is sufficiently high and the popularity distribution of the requested content is not excessively skewed. In addition, motivated by the prospect of bidirectional content exchange, we presented an FD version of the proposed policy, which exhibits a small improvement over the HD version in terms of the delay performance, due to the low probability of bidirectional content exchange. According to the simulation results, increasing this probability does not yield proportional improvements in performance due to the resulting prevalence of the cellular rate over the D2D rate, attributed to offloading. As future work, we plan to generalize the proposed caching scheme to a policy that divides the cacheable content to an arbitrary number of groups and study the impact on performance.

Acknowledgements

This work has been funded by the Research Projects AGAUR DI088-2015, AGAUR 2014-SGR-1551, CellFive (TEC2014-60130-P), and 5Gwireless (641985).



(a) $K = 50$

(b) $K = 100$

(c) $K = 200$

Fig. 9: The 90-th percentile of the content retrieval delay D in terms of the popularity exponent ξ for a) $K = 50$, b) $K = 100$, and c) $K = 200$.

References

[1] C. V. N. Index, "Cisco visual networking index: Global mobile data traffic forecast update, 2015–2020," Cisco, Tech. Rep., 2016.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5g be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[3] B. A. Ramanan, L. M. Drabeck, M. Haner, N. Nithi, T. E. Klein, and C. Sawkar, "Cacheability analysis of http traffic in an operational lte network," in *Wireless Telecommunications Symposium (WTS), 2013*, Apr. 2013, pp. 1–8.

[4] N. Golrezaei, A. F. Molisch, A. G. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 142–149, Apr. 2013.

[5] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5g wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[6] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[7] N. Golrezaei, A. G. Dimakis, and A. F. Molisch, "Scaling behavior for device-to-device communications with distributed caching," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4286–4298, July 2014.

[8] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[9] M. Ji, G. Caire, and A. F. Molisch, "The throughput-outage tradeoff of wireless one-hop caching networks," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6833–6859, Dec. 2015.

[10] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3665–3676, July 2014.

[11] A. Altieri, P. Piantanida, L. R. Vega, and C. G. Galarza, "On fundamental trade-offs of device-to-device communications in large wireless networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 4958–4971, Sept. 2015.

[12] L. Zhang, M. Xiao, G. Wu, and S. Li, "Efficient scheduling and power allocation for d2d-assisted wireless caching networks," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2438–2452, June 2016.

[13] J. Jiang, S. Zhang, B. Li, and B. Li, "Maximized cellular traffic offloading via device-to-device content sharing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 82–91, Jan. 2016.

[14] J. Qiao, X. S. Shen, J. W. Mark, Q. Shen, Y. He, and L. Lei, "Enabling device-to-device communications in millimeter-wave 5g cellular networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 209–215, Jan. 2015.

[15] Y. Niu, C. Gao, Y. Li, L. Su, D. Jin, and A. V. Vasilakos, "Exploiting device-to-device communications in joint scheduling of access and backhaul for mmwave small cells," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2052–2069, Oct. 2015.

[16] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[17] G. R. Maccartney, T. S. Rappaport, S. Sun, and S. Deng, "Indoor office wideband millimeter-wave propagation measurements and channel models at 28 and 73 ghz for ultra-dense 5g wireless networks," *IEEE Access*, vol. 3, pp. 2388–2424, 2015.

[18] M. N. Kulkarni, S. Singh, and J. G. Andrews, "Coverage and rate trends in dense urban mmwave cellular networks," in *2014 IEEE Global Communications Conference*, Dec. 2014, pp. 3809–3814.

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

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

[21] T. Bai, A. Alkhateeb, and R. W. Heath, "Coverage and capacity of millimeter-wave cellular networks," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 70–77, Sept. 2014.

[22] M. Haenggi, *Stochastic Geometry for Wireless Networks*, 1st ed. New York, NY, USA: Cambridge University Press, 2012.

[23] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[24] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 176–189, Jan. 2016.

[25] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: Evidence and implications," in *IN INFOCOM*, 1999, pp. 126–134.

[26] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: A view from the edge, imc," in *In: Proc. of IMC*, 2007.

[27] N. Golrezaei, A. G. Dimakis, and A. F. Molisch, "Wireless device-to-device communications with distributed caching," in *Information Theory*

Proceedings (ISIT), 2012 IEEE International Symposium on, July 2012, pp. 2781–2785.

[28] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 3358–3363.

[29] N. Giatsoglou, K. Ntontin, E. Kartsakli, A. Antonopoulos, and C. Verikoukis, "D2d-aware device caching in mmwave-cellular networks," 2017. [Online]. Available: https://arxiv.org/abs/1703.04935

[30] A. M. Hunter, J. G. Andrews, and S. Weber, "Transmission capacity of ad hoc networks with spatial diversity," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5058–5071, Dec. 2008.

[31] A. K. Gupta, J. G. Andrews, and R. W. Heath, "On the feasibility of sharing spectrum licenses in mmwave cellular systems," *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 3981–3995, Sept. 2016.

[32] Z. Wei, X. Zhu, S. Sun, Y. Huang, L. Dong, and Y. Jiang, "Full-duplex versus half-duplex amplify-and-forward relaying: Which is more energy efficient in 60-ghz dual-hop indoor wireless systems?" *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 12, pp. 2936–2947, Dec. 2015.

[33] L. Li, K. Josiam, and R. Taori, "Feasibility study on full-duplex wireless millimeter-wave systems," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 2769–2773.

[34] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 6727–6740, Dec. 2014.

[35] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1152–1163, June 2014.

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

[37] J.-S. Ferenc and Z. Néda, "On the size distribution of poisson voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518 – 526, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437107007546

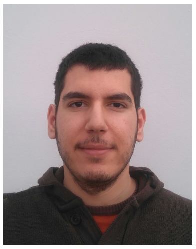
[38] S. M. Yu and S. L. Kim, "Downlink capacity and base station density in cellular networks," in *Modeling Optimization in Mobile, Ad Hoc Wireless Networks (WiOpt), 2013 11th International Symposium on*, May 2013, pp. 119–124.

[39] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[40] F. C. Commission, "Fcc takes steps to facilitate mobile broadband and next generation wireless technologies in spectrum above 24 ghz," FCC, Tech. Rep., 2016.

[41] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2196–2211, Oct. 2015.

[42] A. Thornburg, T. Bai, and R. W. Heath, "Mmwave ad hoc network coverage and capacity," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 1310–1315.



Nikolaos Giatsoglou received the Diploma in Electrical and Computer Engineering from the Aristotle University of Thessaloniki (AUTh), Greece, in 2015, and he is a Ph.D. candidate in the Technical University of Catalonia (UPC). He is also working as a Marie-Curie Early Stage Researcher (ESR) in IQUADRAT Informatica S.L., Barcelona, Spain, in the context of the 5Gwireless european project. His research interests include the performance analysis of wireless protocols, with an emphasis on caching, mmWave, and full-duplex technology.

Konstantinos Ntontin received the Diploma in Electrical and Computer Engineering in 2006, the M.Sc. Degree in Wireless Systems in 2009, and the Ph.D. degree in 2015 from the University of Patras, Greece, the Royal Institute of Technology (KTH), Sweden, and the Technical University of Catalonia (UPC), Spain, respectively. He is the recipient of the 2013 IEEE COMMUNICATIONS LETTERS Exemplary Reviewer Certificate. His research interests are related to the Physical Layer of wireless telecommunications with an emphasis on the performance analysis in fading channels, MIMO systems, array beamforming, and stochastic modeling of wireless channels.

Elli Kartsakli got her Ph.D. degree from the Technical University of Catalonia (UPC) in February 2012 and is currently a senior research engineer in the IQUADRAT R&D team. Her work has been published and presented in multiple journals, magazines, book chapters and international conferences, and she has actively participated in several national and European projects (IAPP-WSN4QoL, ITN-GREENET, IAPP-Coolness, etc.). Her primary research interests include wireless networking, channel access protocols, energy efficient communication protocols, and protocols and architectures for 5G networks and beyond.

Angelos Antonopoulos received his Ph.D. degree from Technical University of Catalonia (UPC) in 2012. He is currently a Researcher in the SMARTECH department of the Technological Telecommunications Centre of Catalonia (CTTC). He has published over 70 peer-reviewed journals, conferences and book chapters on various topics, including energy efficient network planning and sharing, 5G wireless networks, cooperative communications and network economics. He has been nominated as Exemplary Reviewer for the IEEE Communications Letters, and has received the best paper award in IEEE GLOBECOM 2014, the best demo award in IEEE CAMAD 2014, the 1st prize in the IEEE ComSoc Student Competition (as a Mentor) and the EURACON best student paper award in EuCNC 2016.

Christos Verikoukis received his Ph.D. degree from the Technical University of Catalonia (UPC) in 2000. He is currently a Fellow Researcher at CTTC (Head of the SMARTECH department) and an adjunct professor at the Electronics Department of the University of Barcelona (UB). He has published 107 journal papers and over 170 conference papers. He is co-author in 3 books, 16 chapters in different books, and has filed 3 patents. He has supervised 15 Ph.D. students and 5 Post Docs researchers since 2004. He has participated in more than 30 competitive projects and has served as the Principal investigator in national projects in Greece and Spain. He received the best paper award in the IEEE ICC 2011, IEEE ICC 2014, IEEE GLOBECOM 2015, and in the EUCNC 2016 conferences, as well as the EURASIP 2013 Best Paper Award for the Journal on Advances in Signal Processing. He was the general Chair of the 17th, 18th, and 19th IEEE CAMAD, the TPC Co-Chair of the 15th IEEE Healthcom and the 7th IEEE Latincom Conference, and the symposium Co-Chair of the CQRM symposium in the IEEE ICC 2015 & 2016 conference. He is currently the General Co-Chair of the 22th IEEE CAMAD and the CQRM symposium Co-Chair in IEEE Globecom 2017, and the Chair of the IEEE ComSoc Technical Committee on Communication Systems Integration and Modeling (CSIM).