Mobility Management for Heterogeneous Networks: Leveraging Millimeter Wave for Seamless Handover

Omid Semiari[1], Walid Saad[2], Mehdi Bennis[3], and Behrouz Maham[4]

[1]Department of Electrical Engineering, Georgia Southern University, Statesboro, GA, USA, Email: osemiari@georgiasouthern.edu
[2]Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA, Email: walids@vt.edu
[3] Centre for Wireless Communications, University of Oulu, Finland, Email: bennis@ee.oulu.fi
[4]Department of Electrical and Electronic Engineering, Nazarbayev University, Astana, Kazakhstan, Email: behrouz.maham@nu.edu.kz

Abstract—One of the most promising approaches to overcome the uncertainty and dynamic channel variations of millimeter wave (mmW) communications is to deploy dual-mode base stations that integrate both mmW and microwave (μW) frequencies. In particular, if properly designed, such dual-mode base stations can enhance mobility and handover in highly mobile wireless environments. In this paper, a novel approach for analyzing and managing mobility in joint μW-mmW networks is proposed. The proposed approach leverages device-level caching along with the capabilities of dual-mode base stations to minimize handover failures and provide seamless mobility. First, fundamental results on the caching capabilities, including caching probability and cache duration, are derived for the proposed dual-mode network scenario. Second, the average achievable rate of caching is derived for mobile users. Then, the impact of caching on the number of handovers (HOs) and the average handover failure (HOF) is analyzed. The derived analytical results suggest that content caching will reduce the HOF and enhance the mobility management in heterogeneous wireless networks with mmW capabilities. Numerical results corroborate the analytical derivations and show that the proposed solution provides significant reductions in the average HOF, reaching up to 45%, for mobile users moving with relatively high speeds.

I. INTRODUCTION

The proliferation of bandwidth-intensive wireless applications such as social networking, high definition video streaming, and mobile TV have drastically strained the capacity of wireless cellular networks. To cope with this traffic increase, several new technologies are anticipated for 5G cellular systems, including: 1) dense deployment of small cell base stations (SBSs), and 2) leveraging the large amount of available bandwidth at *millimeter wave (mmW)* frequencies [1]. In fact, SBSs will boost the capacity of wireless networks by reducing the cell sizes and removing the coverage holes. Meanwhile, mmW communications will provide high data rates, by exploiting directional antennas and transmitting over a large bandwidth that can reach up to 5 GHz.

However, one of the key practical issues associated with the dense deployment of SBSs is frequent handovers (HOs) which increases the overhead and delay in heterogeneous networks (HetNets). In addition, handover failure (HOF) is more common in HetNets, particularly, for *mobile user equipments (MUEs)* with higher speeds [2]. In fact, due to the small and disparate cell sizes in HetNets, MUEs will not be able to

successfully finish the HO process by the time they trigger HO and pass a target SBS.

To enhance mobility management in HetNets, an extensive body of work has appeared in the literature [3]–[11]. One widely adopted scheme to cope with channel quality variations in mobile environment is the dynamic adaptive streaming over HTTP (DASH) protocol. The authors in [3] analyze the performance of DASH in vehicular networks. In [4], a novel multi-user DASH protocol is proposed to enhance quality of experience for MUEs. However, the DASH protocol offers content segments with different quality (different data rate) and it selects the content with lower quality when throughput decreases which results in a lower quality-of-service (QoS). Therefore, despite its importance, the DASH protocol alone will not be sufficient to meet the stringent requirements of 5G applications such as HD TV, uncompressed video streaming, or virtual reality (VR).

In [5], the authors comprehensively survey mobility management in IP networks. The work presented in [6] overviews existing approaches for vertical handover decisions in HetNets. In [7], the authors analyze the impact of channel fading on mobility management in HetNets. The work in [8] introduces an HO scheme that considers the speed of MUEs to decrease frequent HOs in HetNets. Moreover, the authors in [9] propose an HO scheme that supports soft HO by allowing MUEs to connect with both a macrocell base station (MBS) and SBSs. Furthermore, a distributed mobility management framework is proposed in [10] which uses multiple frequency bands to decouple the data and control planes. In [11], an HO scheme for mmW networks is proposed in which the MBS acts as an anchor for mmW SBSs to manage control signals. Although interesting, the works in [5]–[10] do not consider mmW communications. Moreover, [11] assumes that line-of-sight (LoS) mmW links are always available and provides no analytical results to capture the directional nature of mmW communications.

The main contribution of this paper is a novel mobility management framework that addresses critical handover issues, including frequent HOs and HOF in emerging dense wireless cellular networks with mmW capabilities. In fact, we adopt a model that allows MUEs to cache their requested content by exploiting high capacity mmW connectivity whenever available. Thus, the MUEs will be able to use the cached content and avoid performing any HO, while passing SBSs

with relatively small cell sizes. In addition, we propose a geometric model to derive tractable, closed-form expressions for key performance metrics, including the probability of caching, cumulative distribution function of caching duration, and the average data rate for caching at an MUE over a mmW link. Moreover, we provide insight on the achievable gains for reducing the number of HOs and the average HOF, by leveraging caching in mmW-μW networks. Under practical settings, we show that HOF can be decreased by up to 45%, even for MUEs moving at high speeds.

The rest of this paper is organized as follows. Section II presents the system model. Section III presents the analysis for caching in mobility management. Performance analysis are provided in Section IV. Section V presents the simulation results and Section VI concludes the paper.

II. SYSTEM MODEL

Consider a HetNet composed of an MBS and K SBSs within a set $\mathcal{K}$, distributed uniformly across an area. The SBSs are equipped with mmW front-ends to serve MUEs over either mmW or μW frequency bands [12]. The dual-mode capability allows to integrate mmW and μW radio access technologies (RATs) at the medium access control (MAC) layer of the air interface and reduce the delay and overhead for fast vertical handovers between both RATs. We consider a set $\mathcal{U}$ of U MUEs that are distributed randomly and move across the considered geographical area during a time frame T. Each user $u \in \mathcal{U}$ moves in a random direction $\theta_u \in [0, 2\pi]$, with respect to the $\theta = 0$ horizontal angle, which is assumed fixed for each MUE over a considered time frame T. In addition, we consider that an MUE u moves with an average speed $v_u \in [v_{\min}, v_{\max}]$. The MUEs can receive their requested traffic over either the mmW or μW band.

A. Channel model

The large-scale channel effect over mmW frequencies for a link between an SBS k and an MUE $u \in \mathcal{U}$, in dB, is[1]

$$L(u,k) = 20\log_{10}\left(\frac{4\pi r_0}{\lambda}\right) + 10\alpha \log_{10}\left(\frac{r_{u,k}}{r_0}\right) + \chi, \quad (1)$$

where (1) holds for $r_{u,k} \geq r_0$, with r_0 and $r_{u,k}$ denoting, respectively, the reference distance and distance between the MUE u and SBS k. In addition, α is the path loss exponent, λ is the wavelength at carrier frequency $f_c = 73$ GHz over the E-band, and χ is a Gaussian random variable with zero mean and variance ξ^2. The path loss parameters α and ξ will naturally have different values, depending on whether the mmW link is LoS or non-LoS (NLoS). Over the μW frequency band, the path loss model follows (1), however, with parameters that are specific to sub-6 GHz frequencies.

An illustration of the considered HeNet is shown in Fig. 1. The coverage for each SBS at the μW frequency is shown based on the maximum received signal strength (max-RSS) criteria. In addition, white spaces in Fig. 1 show the areas that are covered solely by the MBS. Here, we observe that
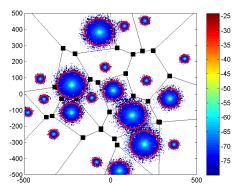
---

[1]The free space path loss model in (1) has been adopted in many existing works, such as in [1] that carry out real-world measurements to characterize mmW large scale channel effects.

Figure 1: SBSs coverage with RSS threshold of -80 dB. Red circles show the simplified cell boundaries.

shadowing effect can adversely increase the ping-pong effect for MUEs. To cope with this issue, the 3GPP standard suggests L1/L3 filtering which basically applies averaging to RSS samples, as detailed in [7].

B. Antenna model and configuration

To overcome the excessive path loss at f_c, the MUEs are equipped with electronically steerable antennas which allow them to achieve beamforming gains at a desired direction. The antenna gain pattern for MUEs follows the simple and widely-adopted sectorized pattern which is given by:

$$G(\theta) = \begin{cases} G_{\max}, & \text{if} \quad \theta < |\theta_m|, \\ G_{\min}, & \text{otherwise}, \end{cases} \quad (2)$$

where θ and θ_m denote, respectively, the azimuth angle and the antennas' main lobe beamwidth. Moreover, $G_{\max}$ and $G_{\min}$ denote, respectively, the antenna gain of the main lobe and side lobes. For MUEs, we use a model similar to the sectorized pattern in (2). However we allow each SBS to form N beams, either by using N antenna arrays or forming multi-beam beamforming. The beam patten configuration of an MBS is shown in Fig. 2, where N equidistant beams in $\theta \in [0, 2\pi]$ are formed. To avoid the complexity and overhead of beam-tracking for mobile users, the location of the SBSs' beams in azimuth is fixed. In fact, an MUE can connect to an SBS over a mmW link, if the MUE traverses the area covered by the mmW beams of this SBS. It is assumed that for a desired link between an SBS k and an MUE u, the overall transmit-receive gain is $\psi_{u,k} = G_{\max}^2$.

C. Traffic model

Video streaming is one of the wireless services with most stringent QoS requirement. The QoS achieved by such services is severely impacted by the delay that results from frequent handovers in HetNets. In addition, HOFs can significantly degrade the performance by making frequent service inter-ruptions. Therefore, our goal is to enhance mobility man-agement for MUEs that request video traffic. Each video content is partitioned into small segments, each of size B bits. The network incorporates caching [13], [14] over the mmW frequency band to store incoming video segments at the MUE's storage, whenever a high capacity mmW connection is available. Considering a cache size of Ψ_u for an arbitrary MUE

u, associated with an SBS k, the number of video segments that can be cached at MUE u will be

$$M^c(u,k) = \min\left\{ \left\lfloor \frac{\bar{R}^c(u,k)t_u^c}{B} \right\rfloor, \frac{\Psi_u}{B} \right\}, \quad (3)$$

where $\lfloor . \rfloor$ and $\min\{.,.\}$ denote, respectively, the floor and minimum operands. In addition, t_u^c is the *caching duration* which is equal to the time needed for an MUE u to traverse a mmW beam at its serving SBS. Considering the small green triangle in Fig. 2 as the current location of an MUE crossing a mmW beam, caching duration will be $t_u^c = r_u^c/v_u$. Moreover, $\bar{R}^c(u,k)$ is the *average achievable rate* for the MUE u during t_u^c. Given $M^c(u,k)$ and the video play rate of Q, specified for each video content, the distance an MUE u can traverse with speed v_u, while playing the cached video content will be

$$d^c(u,k) = \frac{M^c(u,k)}{Q}v_u. \quad (4)$$

In fact, the MUE can traverse a distance $d^c(u,k)$ by using the cached video content without requiring an HO to any of the target cells. Meanwhile, the location information and control signals, such as paging, can be handled by the MBS over the μW frequency band. Here, we note that buffering segments which is conventionally used in sub-6 GHz systems can also be employed within the proposed HO muting scheme. However, by using buffering, the network can only download a limited number of segments in advance. In contrast, with caching, an MUE can store an entire or a large portion of a content. Thus, by using caching and not just buffering, a mobile MUE will be able to traverse a longer distance d^c, while playing the stored content. Furthermore, by using the mmW frequency band, one can leverage the substantial amount of bandwidth which is suitable to realize the idea of fully-fledged content caching in practice.

D. Handover procedure and performance metrics

The HO process in the 3GPP standard proceeds as follows [15]: 1) Each MUE will do a cell search every T_s seconds that can be configured by the network or directly by MUEs, 2) If any target cell offers an RSS plus a hysteresis that is higher than the serving cell, even after L1/L3 filtering of input RSS samples, the MUE will wait for a time-to-trigger (TTT) of ΔT seconds to measure the average RSS from the target cell, 3) If the average RSS is higher than that of the target SBS during TTT, the MUE triggers HO and sends a measurement report to its serving cell. The averaging over the TTT duration will reduce the ping-pong effect resulting from instantaneous channel state information (CSI) variations, and 4) HO will be executed after the serving SBS sends HO information to the target SBS.

In our model, we modify the above HO procedure to leverage the caching capabilities of the MUEs during mobility. Here, we let each MUE dynamically determine T_s, depending on the number of remaining video segments M in the cache, the video play rate, and the MUE's speed. That is, an MUE u will mute the cell search while M/Q is greater than ΔT. In this way, the MUE will have ΔT seconds to search for a target SBS before the cached content runs out.
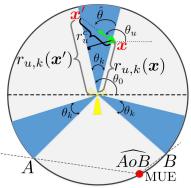

Figure 2: Antenna beam configuration of a dual-mode SBS with $N_k = 3$. Blue shaded areas show the mmW beams.

Next, we consider the HOF as a key performance metric for an HO procedure. One of the main reasons for the potential increase in HOF is due to the relatively small cell sizes in HetNets, compared to the MBS coverage. In fact, HOF occurs if the time-of-stay (ToS) for an MUE is less than the minimum ToS (MTS) required for performing a successful HO. That is,

$$\gamma_{\text{HOF}}(u,k) = \begin{cases} 1, & \text{if } t_{u,k} < t_{\text{MTS}}, \\ 0, & \text{otherwise}, \end{cases} \quad (5)$$

where $t_{u,k}$ is the ToS for MUE u to pass across SBS k coverage. Although a short ToS may not be the only cause for an HOF, it becomes critical for small cell setting and MUEs with high speeds [15]. Next, we adopt a geometric framework to analyze the caching opportunities, in terms of the caching duration t^c, and the average achievable rate $\bar{R}^c$, for MUEs moving at random directions in joint mmW-μW HetNets.

III. ANALYSIS OF MOBILITY MANAGEMENT WITH CACHING CAPABILITIES

First, we investigate the probability of serving an arbitrary MUE over mmW frequencies by a dual-mode SBS.

A. Probability of mmW coverage

In Fig. 2, the small circle shows the intersection of an MUE u's trajectory with the coverage area of an arbitrary SBS k. In this regard, $\mathbb{P}_k^c(N_k, \theta_k)$ represents the probability that MUE u with a random direction θ_u and speed v_u crosses the mmW coverage areas of SBS k. From Fig. 2, we observe that the MUE will pass through the mmW coverage area only if the MUE's direction is inside the angle $\widehat{AoB}$. Hence, we can state the following.

Theorem 1. If the mmW front-end of SBS k has formed $N_k \geq 2$ main lobes, each with a beamwidth $\theta_k > 0$, the probability of mmW coverage will be given by:

$$\mathbb{P}_k^c(N_k, \theta_k) = \left[\frac{N_k\theta_k}{2\pi} \right] + \left[1 - \frac{N_k\theta_k}{2\pi} \right]\left[\frac{1}{2}\left(1 - \frac{1}{N}\right) + \frac{\theta_k}{4\pi} \right]. \quad (6)$$

Proof. See Appendix A. ∎

We can verify (6) by considering an example scenario with $N_k = 3$ and $\theta_k = \frac{2\pi}{3}$. For this example, (6) results in $\mathbb{P}_k^c(N_k, \theta_k) = 1$ which correctly captures the fact that the entire cell is covered by mmW beams.

B. Cumulative distribution function of the caching duration

To enable an MUE to use the cached content while not being associated to an SBS, it is critical to assess the distribution of caching duration t^c for an arbitrary MUE with a random direction and speed. In this regard, consider the small green triangle in Fig. 2, which represents the location of an arbitrary MUE u, $\boldsymbol{x_u} = (x_u, y_u) \in \mathbb{R}^2$, crossing a mmW beam. First, we note that the geometry of a mmW beam for an SBS can be defined by the location of SBS, as well as the sides of the beam angle. Without loss of generality, assume that the SBS of interest is located at the center, such that $\boldsymbol{x_k} = (0, 0)$. Therefore, the two sides of the beam angle will be given by

$$y = x \tan(\theta_0 - \theta_k), y = x \tan(\theta_0), \quad x > 0. \quad (7)$$

Assuming that the MUE u is currently located on the angle side $x = y \cos(\theta_0 - \theta_k)$, as shown by the small triangle in Fig. 2, then θ_0 in (7) will be $\theta_0 = \arccos\left(\frac{x_u}{r_{u,k}(\boldsymbol{x_u})}\right) + \theta_k$, where $r_{u,k}(\boldsymbol{x}) = \sqrt{x_u^2 + y_u^2}$. Hereinafter, we will use the parameter θ_0 to simplify our analysis. Let $F_{t^c}(.)$ be the cumulative distribution function (CDF) of the caching duration t^c. Thus,

$$F_{t_u^c}(t_0) = \mathbb{P}(t_u^c \leq t_0) = \mathbb{P}(r_u^c \leq v_u t_0), \quad (8)$$

where r_u^c denotes the distance that MUE u will traverse across the mmW beam, as shown in Fig. 2. Given the location of MUE $\boldsymbol{x_u}$, the minimum possible distance to traverse, $r_u^{\min}$, is

$$r_u^{\min} = \frac{|x_u \tan \theta_0 - y_u|}{\sqrt{1 + \tan^2 \theta_0}}. \quad (9)$$

In fact, (9) gives the distance of the point $\boldsymbol{x_u}$ from the beam angle side $y = x \tan(\theta_0)$. If $r_u^{\min} > v_u t_0$, then $F_{t_u^c}(t_0) = 0$. Therefore, for the remainder of this analysis we consider $r_u^{\min} \leq v_u t_0$. Next, let $\boldsymbol{x'_u}$ denote the intersection of the MUE's path with line $y = x \tan(\theta_0)$. It is easy to see that $\boldsymbol{x'_u} = (x_u + r_u^c \cos \theta_u, y_u + r_u^c \sin \theta_u)$. Hence, $y_u + r_u^c \sin \theta_u = [x_u + r_u^c \cos \theta_u] \tan \theta_0$, and r_u^c, i.e., the distance that MUE u traverses during the caching duration t^c is given by:

$$r_u^c = v_u t_u^c = \frac{y_u - x_u \tan \theta_0}{\tan \theta_0 \cos \theta_u - \sin \theta_u}. \quad (10)$$

Next, from (8) and (10), the CDF can be written as

$$F_{t_u^c}(t_0) = \mathbb{P}\left(\frac{y_u - x_u \tan \theta_0}{\tan \theta_0 \cos \theta_u - \sin \theta_u} \leq v_u t_0\right). \quad (11)$$

Using the geometry shown in Fig. 2, we find the CDF of caching duration as follows:

Lemma 1. The CDF of caching duration, t^c, for an arbitrary MUE u with speed v_u is given by

$$F_{t^c}(t_0) = \frac{1}{\pi - \theta_k}\left(\arccos\left(\frac{r_u^{\min}}{v_u t_0}\right)\right. \quad (12)$$

$$\left. + \min\left\{\arccos\left(\frac{r_u^{\min}}{r_{u,k}(\boldsymbol{x})}\right), \arccos\left(\frac{r_u^{\min}}{v_u t_0}\right)\right\}\right).$$

Proof. See Appendix B. ∎

The CDF of t^c is shown in Fig. 3 for different MUE distances from the serving SBS. Fig. 3 shows that as the MUE is closer to the SBS, t^c takes smaller values with higher
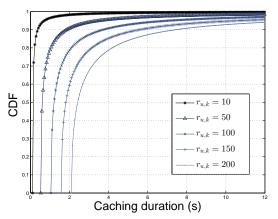

Figure 3: CDF of caching duration t^c.

probability which is expected, since the MUE will traverse a shorter distance to cross the mmW beam.

IV. PERFORMANCE ANALYSIS OF THE PROPOSED CACHING-ENABLED MOBILITY MANAGEMENT SCHEME

Next, we analyze the average achievable rate for content caching, for an MUE with speed v_u, direction θ_u, and initial distance $r_{u,k}(\boldsymbol{x})$ from the serving dual-mode SBS. In addition, we evaluate the impact of caching on mobility management.

A. Average Achievable Rate for Caching

The achievable rate of caching follows

$$R^c(u, k) = \frac{1}{v_u t_u^c} \int_{r_{u,k}(\boldsymbol{x})}^{r_{u,k}(\boldsymbol{x'})} w \log\left(1 + \frac{\beta P_t \psi r_{u,k}^{-\alpha}}{wN_0}\right) dr_{u,k}, \quad (13)$$

where $\beta = (\frac{\lambda}{4\pi r_0})^2 r_0^\alpha$. The integral in (13) is taken over the line with length r_u^c that connects the MUE location $\boldsymbol{x}$ to $\boldsymbol{x'}$, as shown in Fig. 2.

Theorem 2. The average achievable rate for an MUE u served by an SBS k, $\bar{R}^c(u, k)$, is given by

$$\bar{R}^c(u, k) = \mathbb{P}_k^c(N_k, \theta_k) R^c(u, k), \quad (14)$$

$$= \delta_2 \int_{f(\theta_k)}^{f(0)} \frac{1}{f^2(\theta)} \log\left(1 + \delta_1 f^\alpha(\theta)\right) df(\theta), \quad (15)$$

$$\stackrel{(a)}{=} \frac{\delta_2}{\ln(2)}\left[2\sqrt{\delta_1} \arctan(\sqrt{\delta_1} f(\theta_k)) - \frac{\ln(\delta_1 f^2(\theta_k) + 1)}{f(\theta_k)},\right.$$

$$\left. -2\sqrt{\delta_1} \arctan(\sqrt{\delta_1} f(0)) + \frac{\ln(\delta_1 f^2(0) + 1)}{f(0)}\right], \quad (16)$$

where $\delta_1 = \beta P_t \psi (r_{u,k}(\boldsymbol{x}) \sin \hat{\theta})^{-\alpha}/wN_0$. Moreover, $\delta_2 = wr_{u,k}(\boldsymbol{x}) \sin \hat{\theta} \mathbb{P}_k^c(N_k, \theta_k)/v_u t^c$, and $\hat{\theta} = \theta_u - \theta_0 + \theta_k$. For (a) to hold, we set $\alpha = 2$ which is a typical value for the path loss exponent of LoS mmW links [1].

Proof. Theorem 1 implies that with probability $1 - \mathbb{P}_k^c(N_k, \theta_k)$, only μW coverage is available for an MUE. Therefore, average achievable rate for caching over the mmW frequencies is given by (14). To simplify (14), we have

$$r_{u,k} \cos \theta = r_{u,k}(\boldsymbol{x}) + r_u \cos \hat{\theta}, \ r_{u,k} \sin \theta = r_u \sin \hat{\theta}, \quad (17)$$

where $\hat{\theta} = \theta_u - \theta_0 + \theta_k$ and θ is an angle between the line connecting MUE to SBS, ranging from 0 to θ_k. Moreover, r_u

is the current traversed distance, with $r_u = r_u^c$ once the MUE reaches $\boldsymbol{x}'$ by the end of caching duration, as shown in Fig. 2. From (17), we find $r_{u,k} = r_{u,k}(\boldsymbol{x}) \sin\hat\theta / \sin(\hat\theta - \theta)$. By changing the integral variable r_u to θ, we can write (14) as

$$\bar{R}^c(u,k) = \delta_2 \int_0^{\theta_k} \log\left(1 + \delta_1 \sin^\alpha(\hat\theta - \theta)\right) \frac{\cos(\hat\theta - \theta)}{\sin^2(\hat\theta - \theta)} d\theta, \quad (18)$$

where $\delta_1 = \beta P_t \psi (r_{u,k}(\boldsymbol{x}) \sin\hat\theta)^{-\alpha} / w N_0$ and $\delta_2 = w r_{u,k}(\boldsymbol{x}) \sin\hat\theta \mathbb{P}_k^c(N_k, \theta_k) / v_u t^c$. Next, we can directly conclude (15) from (18) by substituting $f(\theta) = \sin(\hat\theta - \theta)$ in (18). For $\alpha = 2$, which is a typical value for the path loss exponent for LoS mmW links, (15) can be simplified into (16) by taking the integration by parts in (15). ∎

B. Achievable gains of caching for mobility management

From (3), (4), and (16), we can find $d^c(u,k)$ which is the distance that MUE u can traverse, while using the cached video content. On the other hand, by having the average inter-cell distances in a HetNet, we can approximate the number of SBSs that an MUE can pass over distance $d^c(u,k)$. Hence, the average number of SBSs that MUE is able to traverse without performing cell search for HO is

$$\zeta \approx \left\lfloor \frac{\mathbb{E}\left[d^c(u,k)\right]}{l} \right\rfloor, \quad (19)$$

where the expected value is taken, since $d^c(u,k)$ is a random variable that depends on θ_u. Moreover, l denotes the average inter-cell distance. Here, we note that

$$\mathbb{E}\left[d^c(u,k)\right] = \int_0^\infty (1 - F_{t_u^c}(v_u t)) dt, \quad (20)$$

where $F_{t^c}(.)$ is derived in Lemma 1. We note that (20) is the direct result of writing an expected value in terms of CDF. Based on the definition of ζ in (19), we can conclude the following.

Remark 1. The proposed caching scheme will reduce the average number of HOs by $1/\zeta$ factor.

Furthermore, from the definition of HOF γ_{HOF} in (5), we can define the probability of HOF as $\mathbb{P}(D_{u,k} < v_u t_{\text{MTS}})$ [16], where $D_{u,k} = t_{u,k} / v_u$, and $t_{u,k}$ is the ToS. To compute the HOF probability, we use the probability density function (PDF) of a random chord length within a circle with radius a, as follows:

$$f_D(D) = \frac{2}{\pi\sqrt{4a^2 - D^2}}, \quad (21)$$

where (21) relies on the assumption that one side of the chord is fixed and the other side is determined by choosing a random $\theta \in [0, \pi]$. This assumption is in line with our analysis as shown in Fig. 2. Using (21), we can find the probability of HOF as follows:

$$\mathbb{P}(D_{u,k} < v_u t_{\text{MTS}}) = \int_0^{v_u t_{\text{MTS}}} \frac{2}{\pi\sqrt{4a^2 - D^2}} dD, \quad (22)$$

$$= \frac{2}{\pi} \arcsin\left(\frac{v_u t_{\text{MTS}}}{2a}\right). \quad (23)$$

In fact, γ_{HOF} is a Bernoulli random variable with a probability of success that depends on the MUE's speed, cell radius, and t_{MTS}. Hence, by reducing the number of HOs by $1/\zeta$ factor,

Table I: Simulation parameters

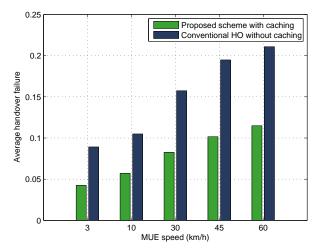| Notation	Parameter	Value
f_c	Carrier frequency	73 GHz
$P_{t,k}$	Total transmit power of SBSs	$[20, 27, 30]$ dBm
K	Total number of SBSs	50
w	Available Bandwidth	5 GHz
$(\alpha_{\text{LoS}}, \alpha_{\text{NLoS}})$	Path loss exponent	$(2, 3.5)$ [1]
d_0	Path loss reference distance	1 m [1]
$G_{\max}$	Antenna main lobe gain	18 dB
$G_{\min}$	Antenna side lobe gain	-2 dB
N_k	Number of mmW beams	3
θ_m, θ_k	beam width	10°
N_0	Noise power spectral density	-174 dBm/Hz
t_{MTS}	Minimum time-of-stay	1s [15]
Q	Play rate	1k segments per second
B	Size of video segments	1 Mbits
v_u	MUE speed	$[3, 10, 30, 45, 60]$ km/h

Figure 4: HOF vs different MUE speeds.

the proposed scheme will reduce the expected value of the sum $\sum \gamma_{\text{HOF}}$, taken over all SBSs that an MUE visits during the considered time T.

V. SIMULATION RESULTS

For simulations, we consider a HetNet composed of $K = 50$ SBSs distributed uniformly across a circular area with radius 500 meters with the MBS located at the center and a minimum inter-cell distance of 30 meters. The main parameters are summarized in Table I. In our simulations, we consider the overall transmit-receive antenna gain from an interference link to be random. All statistical results are averaged over a large number of independent runs.

Fig. 4 compares the average HOF of the proposed scheme with a conventional HO mechanism that relies on the average RSS to perform HO and does not exploit caching. The results clearly demonstrate that caching capabilities, as proposed here, will significantly improve the HO process for dense HetNets. In fact, the results in Fig. 4 show that caching over mmW frequencies will reduce HOF for all speed, reaching up to 45% for MUEs with $v_u = 60$ km/h.

Furthermore, Fig. 5 shows the achievable rate of caching for an MUE with $v_u = 60$ km/h, as function of different initial distances $r_{u,k}(\boldsymbol{x})$ for various θ_u. The results in Fig. 5 show that even for MUEs with high speeds, the achievable rate of caching is significant, exceeding 10 Gbps, for all θ_u values and inital distance of 20 meters from the SBS. However,
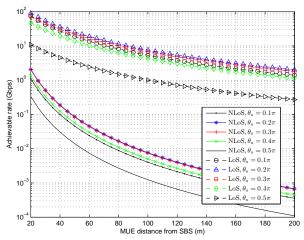
Figure 5: Achievable rate of caching vs $r_{u,k}(\boldsymbol{x})$ for different θ_u.

we can observe that the blockage can noticeably degrade the performance. In fact, for NLoS scenarios, the maximum achievable rate at the distance of 20 meters reduces to 2 Gbps.

VI. CONCLUSIONS

In this paper, we have proposed a comprehensive framework for mobility management in integrated microwave-millimeter wave networks. In particular, we have shown that by smartly caching video contents while exploiting the dual-mode nature of the network's base stations, one can provide seamless mobility to the users. We have derived various fundamental results on the probability and the achievable rate for caching video contents by leveraging millimeter wave high capacity transmissions. We have shown that caching provides significant gains in reducing the number of handovers. Numerical results have corroborated our analytical results and showed that the significant rates for caching can be achieved over the mmW frequencies, even for fast mobile users. The results also have shown that the proposed approach substantially decreases the handover failures in heterogeneous networks.

APPENDIX A

Due to the equidistant beams, we have

$$\widehat{AoB} = \frac{1}{2}AB = \frac{1}{2}\left[2\pi - AoB\right] \qquad (24)$$
$$= \frac{1}{2}\left[2\pi - \left(\frac{2\pi}{N_k} - \theta_k\right)\right] = \left(1 - \frac{1}{N_k}\right)\pi + \frac{\theta_k}{2}.$$

Given that an arbitrary MUE can enter the circle in Fig. 2 with any direction, it will be instantly covered by mmW with probability $\mathbb{P}(\boldsymbol{x}_u \in \mathcal{A}) = \frac{N_k\theta_k}{2\pi}$, where $\mathcal{A} \subset \mathbb{R}^2$ denotes the part of circle's perimeter that overlaps with mmW beams. Thus,

$$\mathbb{P}_k^c(N_k, \theta_k) = \mathbb{P}(\boldsymbol{x}_u \in \mathcal{A}) + \left[1 - \mathbb{P}(\boldsymbol{x}_u \in \mathcal{A})\right]\frac{1}{2\pi}\widehat{AoB}, \quad (25)$$

where (25) is resulted from the fact that $\theta_u \sim U[0, 2\pi]$. Therefore, from (24) and (25), the probability of crossing a mmW beam follows (6).

APPENDIX B

From (8), $F_{t^c}(t_0) = \mathbb{P}(r_u^c \leq v_u t_0)$. To find this probability, we note that $r_u^c \leq v_u t_0$ if MUE moves between two line segments of length $v_u t_0$ that connect MUE to line

$y = x\cos\theta_0$. Depending on $r_{u,k}(\boldsymbol{x})$, the intersection of line segment with $y = x\cos\theta_0$ may have one or two solutions. In case of two intersection points, the two line segments will make two equal angles with the perpendicular line from $\boldsymbol{x}_u$, to $y = x\cos\theta_0$, which each is obviously equal to $\pi - (\pi/2 - \theta_k) - \hat{\theta} = \pi/2 + \theta_k - \hat{\theta} = \arccos\left(\frac{r_u^{\min}}{v_u t_0}\right)$. Therefore,

$$F_{t^c}(t_0) = \frac{2}{\pi - \theta_k}\arccos\left(\frac{r_u^{\min}}{v_u t_0}\right). \qquad (26)$$

In fact, θ_u must be within a range of $\pi - \theta_k$ for $r_u^c \leq v_u t_0$ to be valid. Now, if this angle is greater than $\pi/2 - \theta_k$, only one intersection point exists. Equivalently,

$$F_{t^c}(t_0) = \frac{1}{\pi - \theta_k}\left(\arccos\left(\frac{r_u^{\min}}{v_u t_0}\right) + \arccos\left(\frac{r_u^{\min}}{r_{u,k}(\boldsymbol{x})}\right)\right). \quad (27)$$

Integrating (26) and (27), the CDF for caching duration can be written as (12).

REFERENCES

[1] A. Ghosh, R. Ratasuk, P. Moorut, T. S. Rappaport, and S. Sun, "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.

[2] D. Lopez-Perez, I. Guvenc, and X. Chu, "Mobility management challenges in 3GPP heterogeneous networks," *IEEE Communications Magazine*, vol. 50, pp. 70–78, December 2012.

[3] C. Muller, S. Lederer, and C. Timmerer, "An evaluation of dynamic adaptive streaming over http in vehicular environments," in *Proc. of the 4th Workshop on Mobile Video*, pp. 37–42, February 2012.

[4] A. E. Essaili, D. Schroeder, D. Staehle, M. Shehada, W. Kellerer, and E. Steinbach, "Quality-of-experience driven adaptive http media delivery," in *2013 IEEE International Conference on Communications (ICC)*, pp. 2480–2485, June 2013.

[5] I. F. Akyildiz, J. Xie, and S. Mohanty, "A survey of mobility management in next-generation all-IP-based wireless systems," *IEEE Wireless Communications*, vol. 11, pp. 16–28, Aug 2004.

[6] A. Ahmed, L. M. Boulahia, and D. Gaiti, "Enabling vertical handover decisions in heterogeneous wireless networks: A state-of-the-art and a classification," *IEEE Communications Surveys Tutorials*, vol. 16, pp. 776–811, Second 2014.

[7] K. Vasudeva, M. Simsek, D. Lopez-Perez, and I. Guvenc, "Impact of channel fading on mobility management in heterogeneous networks," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, pp. 2206–2211, June 2015.

[8] M. Khan and K. Han, "An optimized network selection and handover triggering scheme for heterogeneous self-organized wireless networks," *Mathematical Problems in Engineering*, vol. 16, pp. 1–11, 2014.

[9] H. Zhang, N. Meng, Y. Liu, and X. Zhang, "Performance evaluation for local anchor-based dual connectivity in 5G user-centric network," *IEEE Access*, vol. 4, pp. 5721–5729, Sept. 2016.

[10] I. Elgendi, K. S. Munasinghe, and A. Jamalipour, "Mobility management in three-tier sdn architecture for densenets," in *2016 IEEE Wireless Communications and Networking Conference*, pp. 1–6, April 2016.

[11] S. G. Park and Y. S. Choi, "Mobility enhancement in centralized mmwave-based multi-spot beam cellular system," in *Proc. of International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 200–205, Oct 2015.

[12] O. Semiari, W. Saad, and M. Bennis, "Joint millimeter wave and microwave resources allocation in cellular networks with dual-mode base stations," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 4802–4816, July 2017.

[13] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 1046–1061, May 2017.

[14] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 3949–3963, June 2016.

[15] 3GPP, "E-UTRA: Mobility enhancements in heterogeneous networks," *3rd Generation Partnership Project*, vol. Release 11, Sept. 2012.

[16] C. H. M. de Lima, M. Bennis, and M. Latva-aho, "Modeling and analysis of handover failure probability in small cell networks," in *Proc. of IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 736–741, April 2014.