Mobility Management in Cache-Enabled Millimeter Wave Networks

Walid Saad

Electrical and Computer Engineering Department, Network Science, Wireless, and Security (NetSciWiS) Group

Joint work with O. Semiari et al.

Email: walids@vt.edu
Group: http://www.netsciwiwis.com
Personal: http://resume.walid-saad.com
Outline

- Overview on mobility management in 5G
- Proposed mobility management framework
  - Dual-mode base stations
  - System model
  - Mobility management with caching capabilities
  - Performance analysis
  - Dynamic resource management with matching
- Simulation results
- Conclusions
Challenges of mobility management in 5G

- Frequent handover (HO) due to large number of small cells
- HO failure (HOF) is probable, due to reduced cell sizes: users will not be able to successfully finish the HO process by the time they trigger HO and pass a target small cell
- Substantial data rates are required to support emerging technologies such as vehicle-to-everything (V2X) communications.
- HO imposes large control overhead, reducing effective resources for data transmissions.
- With the concept of driverless cars, cellular networks must manage more number of mobile users moving with high speeds.
Mobility management in 5G

Prior results:

- Control and data traffic are managed at microwave bands.
- High speed users are managed mainly by macro base stations.
- Leveraging mmW technology is typically undermined (due to complexity for beam-training, etc) in mobile scenarios.
- Dynamic adaptive streaming over HTTP (DASH) protocol is robust against link quality variations at the expense of reduced content quality. Thus, it will not alone be sufficient to satisfy QoS requirements of many 5G applications.

Our solution approach:

- Enabling control and data planes separation in dual-mode BSs with both millimeter wave (mmW) and microwave radio access technologies (RATs) and caching capabilities
Dual-mode base stations enabled with caching

- Enabling tight integration of mmW and microwave RATs at the MAC layer:
  - Fast switching between RATs with minimum latency (critical for mmW)
  - mmW RAT can be used instantaneously (when available) to cache a large amount of data at the user device, even in high-speed scenarios
  - MUEs will use the cached content and avoid performing any HO, while passing SBSs.
  - Location and paging information can be handled by the microwave RAT.
System Model

- Heterogeneous networks with small cells distributed uniformly
- $K$ Dual-mode small cells
- $U$ mobile user equipment (MUEs) moving at random directions $\theta_u \in [0, 2\pi]$
- Each MUE $u$ moves with an average speed $v_u \in [v_{\text{min}}, v_{\text{max}}]$
- Caching is enabled at mmW RAT

[Diagram showing approximated cell boundary and Voronoi partition]
System Model

- Channel model:
  \[ L(u, k) = 20 \log_{10} \left( \frac{4\pi r_0}{\lambda} \right) + 10\alpha \log_{10} \left( \frac{r_{uk}}{r_0} \right) + \gamma, \]
  - large-scale channel effect
  - wavelength
  - reference distance
  - MUE-BS distance

- Antenna gain pattern for MUEs at each cell sector
  \[ G(\theta) = \begin{cases} 
  G_{\text{max}}, & \text{if } \theta < |\theta_{\text{m}}|, \\
  G_{\text{min}}, & \text{otherwise}, 
\end{cases} \]
  - Gain of main lobe
  - Gain of sidelobes
  - Main lobe beamwidth
  - Shadowing effect
Traffic model:
- Online video streaming is considered.
- Each video content is partitioned into small segments, each of size $B$ bits.
- Video segments can be cached at the MUE whenever a high-capacity mmW link is available.

System Model

\[ M^c(u, k) = \min \left\{ \frac{\bar{R}^c(u, k) t^c_u}{B}, \frac{\Psi_u}{B} \right\}, \]

- Number of cached video segments
- Maximum cache size
- Average achievable data rate
- Caching duration: time needed for an MUE to traverse a mmW beam
Mobility management with caching capabilities

- We develop geometric analysis to find:
  - Probability of mmW coverage
  - Caching duration
  - Average achievable data rate for caching

Distance that MUE traverses across the mmW beam
Mobility management with caching capabilities

- **Theorem 1.** Probability of mmW coverage will be given by:

\[
P_k^c(N_k, \theta_k) = \left[ \frac{N_k \theta_k}{2\pi} \right] + \left[ 1 - \frac{N_k \theta_k}{2\pi} \right] \left[ \frac{1}{2} \left( 1 - \frac{1}{N_k} \right) + \frac{\theta_k}{4\pi} \right]
\]

Number of mmW antenna sectors (must be greater than 2). It is equal to 3 in the shown example

- **Proof:**
  - Proof follows geometry and the fact that direction of MUE is chosen uniformly and randomly between 0 and 2\(\pi\). 

Beamwidth of the main lobe.
Mobility management with caching capabilities

- **Lemma 1.** The CDF of caching duration is:

\[
F_{tc}(t_0) = \frac{1}{\pi - \theta_k} \left( \arccos \left( \frac{r_{u \min}}{v_u t_0} \right) + \min \left\{ \arccos \left( \frac{r_{u \min}}{r_{u,k}(x)} \right), \arccos \left( \frac{r_{u \min}}{v_u t_0} \right) \right\} \right)
\]

- Minimun distance to traverse across mmW beam
- BS-MUE distance at location \(x\)

Caching duration takes smaller values with higher probability as BS-MUE distance decreases
The average achievable rate for an MUE served by an SBS is given by:

\[
\bar{R}^c(u, k) = \mathbb{P}_k^c(N_k, \theta_k) \mathcal{R}^c(u, k),
\]

\[
= \delta_2 \int_{f(\theta_k)} f^{(0)} \frac{1}{f^2(\theta)} \log (1 + \delta_1 f^\alpha(\theta)) \, df(\theta),
\]

\[
= \delta_2 \left[ 2 \sqrt{\delta_1} \arctan(\sqrt{\delta_1} f(\theta_k)) - \frac{\ln(\delta_1 f^2(\theta_k) + 1)}{f(\theta_k)} \right]
\]

\[
- 2 \sqrt{\delta_1} \arctan(\sqrt{\delta_1} f(0)) + \frac{\ln(\delta_1 f^2(0) + 1)}{f(0)}
\]

Function of mmW coverage probability

Achievable caching data rate

Function of SNR

\[
f(\theta) = \sin(\hat{\theta} - \theta)
\]
Performance Analysis
(probablility of handover failure)

- Handover failure (HOF):
  \[
  \gamma_{\text{HOF}}(u, k) = \begin{cases} 
  1, & \text{if } t_{u,k} < t_{\text{MTS}}, \\
  0, & \text{otherwise},
  \end{cases}
  \]
  
  HOF is modeled as a Bernoulli random variable.

  Minimum time-of-stay
  time-of-stay of MUE \( u \) traversing BS \( k \)

  Probability of HOF (success probability)

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

  \[
  D_{u,k} = \frac{t_{u,k}}{v_u}
  \]

  Cell radius

  \[
  = \frac{2}{\pi} \arcsin \left( \frac{v_u t_{\text{MTS}}}{2a} \right).
  \]
Performance Analysis (Reducing HOF)

Average distance that the MUE can traverse using the cached content:

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

average number of small cells that MUE is able to traverse without performing cell search for HO is:

$$\zeta \approx \frac{\mathbb{E} [d^c(u, k)]}{l},$$

Average inter-cell distance.

- By reducing the number of HOs by $1/\zeta$, the proposed scheme will reduce the expected value of the sum $\sum \gamma_{HOF}$ taken over all small cells an MUE visits during the considered time period $T$. 
Simulation Results

- BSs are uniformly distributed across a circular area with radius 500 meters.

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

- Average caching data rate for an MUE with speed 60 km/h
- Achievable rate can be significant, exceeding 10 Gbps
- Blockage can significantly degrade the performance
Simulation Results

- Proposed scheme is compared with existing HO approaches that do not employ caching and HO muting.

Up to 45% reduction in HOF by using the proposed caching-enabled mobility management scheme.
MUEs have a flexibility to perform either a vertical or horizontal HO, while moving to their chosen target cell.

Caching enables MUEs to skip a certain HO, depending on the cache size.

Accordingly, there are three HO actions possible for an arbitrary MUE that is being served by an SBS:

1) Execute an HO for a new assignment with a target SBS,
2) Use the cached content and mute HO,
3) Perform an HO to the MBS.
Cache-Enabled Mobility Management

Minimize the number of MUEs associated with MBS $k_0$.

Probability of HOF must be Less than a threshold (reliability).

Ensures there will be enough cached Video segments if the MUE is not Assigned to any SBS or the MBS.

*Goal: Maximize HOs to SBSs for enhanced offload*
Cache-Enabled Mobility Management

- Problem is NP hard and challenging to solve.
- Centralized solutions are not scalable.
- They incur significant latency thus are not suitable for real-time applications such as online video streaming.
- These solutions will typically rely on the current network instances, such as the location, speed and cache size of the MUEs, and, hence, they fail to capture the dynamics of the system.
Clearly:

- Taking into account future network information, such as the estimated distance from the next target SBS, is imperative to effectively maximize the traffic offloads from the MBS.
Mobility Management with Dynamic Matching Theory

We propose a solution based on *dynamic matching theory* that is both tractable and dynamics-aware.

In wireless networks: *resource allocation, cell association*
Utility of MUEs and SBSs

\[ \Phi(u, k) = P_u^{th} - \mathbb{P} \left( \sum_{k \in \mathcal{K}} \zeta(u, k) D_{u,k} < v_u t_{MTS} \right) \]

\[ = P_u^{th} - \frac{2}{\pi} \arcsin \left( \frac{v_u t_{MTS}}{2a_k} \right). \]

Utility that MUE $u$ assigns to SBS $k$

Utility increases with cell radius $a_k$

Utility that SBS $k$ assigns to MUE $u$

Utility is larger for MUEs that are not capable of using caching for the next $T_s$ time duration.

\[ \Gamma(u, k) = T_s - \frac{Q_u}{Q}. \]
Preference Profile of an MUE

\[ k \succ_u k' \iff \Phi(u, k) > \Phi(u, k'), \]

SBS \( k \) is strictly more preferred than SBS \( k' \)

\[ u \succ_u k \iff \Phi(u, k) < 0, \]

Individual rationality constraint
Preference Profile of an SBS

\[ u \succeq_k u' \iff \Gamma(u, k) > \Gamma(u', k), \]

MUE \( u \) is strictly more preferred than MUE \( u' \)

\[ k \succeq_k u \iff \Gamma(u, k) < 0, \]

Individual rationality constraint
Matching Game and Stability

- The proposed matching game is formally defined as a tuple

\[ \Pi \triangleq (\mathcal{U} \cup \mathcal{K}, \succ_u, \succ_k) \]

Where \( \succ_u = \{ \succ_u \}_{u \in \mathcal{U}} \) and \( \succ_k = \{ \succ_k \}_{k \in \mathcal{K}} \).

- One desirable solution concept is to find a *two-sided stable matching* between the MUEs and SBSs.
Dynamic matching for mobility management in HetNets

- In conventional single-stage matching, the preference profiles of the MUEs and SBSs only depend on the current state of the system, such as the location of MUEs, and the cache sizes.

- The deferred acceptance algorithm cannot guarantee stability, if the preference of the MUEs change after HOFs.

- To achieve stability for dynamic settings, such as in Scenarios 1&2, we need to incorporate the post-HO scenarios into the matching game, such that no MUE can block the stability even after experiencing an HOF.
Dynamic matching for mobility management in HetNets

- To account for possible scenarios that may occur after HO, we consider a *two-stage dynamic matching game* that incorporates within the preference profiles, some of the possible scenarios that may face the MUEs and base stations after handover execution.

Dynamic Matching: preference profiles over different *association plans* rather than SBSs
Dynamic matching for mobility management in HetNets

- An association plan is defined as a sequence of two matchings for a given MUE or SBS.

- For example, \( kk' \) is an association plan that indicates an MUE will be assigned to the SBS \( k \) followed by another HO to SBS \( k' \).

- Two-stage dynamic matching:

\[
\mu^\dagger : \mathcal{U} \cup \mathcal{K}' \to (\mathcal{U} \cup \mathcal{K'})^2, \text{ such that } \mu^\dagger(u) = (\mu_1(u), \mu_2(u)),
\]
Dynamic matching for mobility management in HetNets

- In a dynamic matching problem, one looks at two-stage stability notions
  - Ex-ante stability is defined as a period-1 blocking pair
  - Either the MUEs or the BSs may block the matching, after knowing the outcome of the first matching.
  - Dynamically stable matching is defined as a two-period blocking
- Any dynamically stable matching is also an ex ante stable matching. However, ex ante stability does not guarantee dynamic stability.
- We develop an algorithm that finds a dynamically stable solution for the proposed mobility management problem.
Proposed Algorithm (phase 1)

Algorithm 2 Proposed Algorithm for Dynamic Matching Between MUEs and BSs

Inputs: Preference plans $\kappa$ for all MUEs, MBS, and SBSs.
Outputs: Dynamically stable matching $\mu^*$.

Phase 1:
1: For each MUE $u \in \mathcal{U}$, if $uu \succ_u \kappa$, for all $\kappa \in \mathcal{P}_u$, then $u$ does not send any plan proposal to the BSs. Otherwise, MUE $u$ sends a plan proposal to a BS, according to the most preferred plan $\kappa_u^*$.
2: Each SBS $k \in \mathcal{K}$ receives the plan proposals and tentatively accepts most preferred plans (also compared to plans that are previously accepted), such that the quota $U_k^n$ is not violated at each period. Clearly, any accepted plan $\kappa$ by SBS $k$ satisfies $\kappa \succeq_k kk$.
3: repeat Steps 1 to 2
4: until No plan is rejected. The yielded ex ante stable matching is denoted by $\mu^\dagger = (\mu_1^\dagger, \mu_2^\dagger)$.
Proposed Algorithm (phase 2)

### Algorithm 2 Proposed Algorithm for Dynamic Matching Between MUEs and BSs

**Inputs:** Preference plans $\kappa$ for all MUEs, MBS, and SBSs.

**Outputs:** Dynamically stable matching $\mu^*$.  

**Phase 2:**

5: if $\exists u \in \mathcal{U}, \mu_2^+(u) = u$, then apply DA algorithm in Algorithm 1 to the subset of MUEs with $\mu_2^+(u) = u$ and the subset of BSs with $|\mu_2^+(k)| < U_{k}^{th}$, considering the constraints in (27) and (28). Return yielded matching.

6: else
7: return $\mu^+$.
8: end if

Phase 2 resolves any possible period-2 blocking cases.
Proposed Algorithm (phase 2)

- **Proposition 1.** Phase 1 of the proposed algorithm in Algorithm 2 converges to an ex ante stable association between MUEs and BSs.

- **Theorem 3.** The proposed two-stage algorithm in Algorithm 2 is guaranteed to converge to a dynamically stable association between MUEs and BSs.
Simulation Results

- Load of the target SBS vs. the number of MUEs.

Load of the target cell reduces up to 45% when speed increases from 8 to 10 m/s for 40 MUEs.
Simulation Results

- Energy savings for inter-frequency measurements vs number of MUEs.

The proposed scheme achieves up to 80%, 52%, and 29% gains in saving energy, respectively, for MUE speeds $v_u = 8; 10; \text{ and } 12 \text{ m/s}$ by leveraging cached segments and muting unnecessary cell search (150 mJ is the reference for conventional HO).
Simulation Results

- Signaling overhead vs. number of MUEs.

Although mobility management is, in general, more challenging for high speed MUEs, the overhead of the proposed algorithm decreases for high speed scenarios.
Conclusions

- Fundamental results on the caching capabilities, including caching probability, duration, and the average achievable rate of caching are derived for mobile users.
- Impact of caching on the number of handovers and the average handover failure is analyzed.
- The results show that the proposed mobility management framework yields significant performance gains, in terms of reducing the number of handover failures.

Finally....

Thank You Questions?