

Offloading capability of D2D communications on moving nodes

Antonia Maria Masucci*, Salah Eddine Elayoubi†,

*Orange Labs

44 Avenue de la République, 92320 Chatillon, France

Email: antoniamaria.masucci@orange.com

†CentraleSupélec, L2S

3 Rue Joliot Curie, 91190 Gif-sur-Yvette

Email: salaheddine.elayoubi@centralesupelec.fr

Abstract—We provide a flow level analysis of the offloading capability of D2D technology when devices endowed by caches are moving in the network, within the cell and among cells. We introduce the mobility with respect to the receiving user when we consider D2D communication and with respect to the base station when we consider the classical cellular communication. Due to mobility, the quality of both D2D and cellular communication may change over time which makes the analysis challenging. We describe the system through an exact Markovian model, however, since it does not scale with the number of admissible users we provide a closed form approximation model based on multi-class Processor Sharing queue. Numerical results validate the accuracy of our approximation and show that mobility provides high capacity gains and a large offloading potential of D2D when devices are equipped with caches.

I. INTRODUCTION

Mobile data traffic is growing extensively, increasing load on both radio and backhaul interfaces. In order to cope with this traffic increase while meeting the stringent Quality of Service (QoS) expectations of users, operators deploy small cells near traffic hotspots that alleviate load on the radio interface, but induce additional backhaul deployment costs. In order to reduce backhaul load, the authors of [1] propose a combination of dense small cells deployment with local content caching at small cells. As caches are populated during low load hours, this solution alleviates the load on the backhaul during peak hours. It however still requires the deployment of wired backhaul and induces CAPEX and OPEX costs at small cell sites. A more flexible small cell solution is the deployment of small cells on top of moving vehicles (e.g. buses) with wireless backhaul. A Processor Sharing (PS) analysis in [2] has shown that this is an efficient solution to offload traffic hotspots in the congested cells.

With the advent of Device-to-Device (D2D) communications, direct communication among devices bypassing traditional cellular networks. a new cheap and flexible network densification strategy is possible, exploiting the devices as moving small cells. In order to cope with the wireless backhaul issues, devices are also used as local caches that store popular contents and serve them to nearby users. Optimal caching strategies for this scheme have been derived in [3] and an offloading gain up to 35% have been shown. However, this

analysis neglects the impact of mobility and considers that both caches and users are static. Authors in [4] and [5] tried to optimize the caching strategy in devices depending on their mobility profiles. They found out that devices moving with low or high speeds have to cache the most popular files, the former to satisfy their own needs and the latter for exploiting the big number of contacts they have with other users. Users with medium velocity cache less popular files to avoid the duplication. A full buffer traffic, where a constant number of users download constantly files, is considered and the percentage of files downloaded via D2D is investigated. A 5G architecture suitable for mass events is analyzed in [6] where wearables communicate with gateways and aerial access points provide aid in offloading the cellular network via D2D link management. In [7], the authors analyze the effects of user mobility on two applications of D2D communication, content caching and computation offloading, and design an incentive mechanism for users to participate.

However, a limitation of these works is the lack of performance models that consider mobility of caches and users when retrieving a file using D2D. We propose in this work a model that analyzes the D2D offloading capability when devices, endowed by a cache, are moving inside the cell and between cells. When a user requests a content that is in the cache of a neighboring device, it will be downloaded through D2D communication as long as the transmitter is in its proximity. If the transmitter becomes out of reach before the content is completely retrieved, the download is resumed from another neighboring device that has the content, or directly from the base station. Mobility is considered between devices and with respect to the base station and it is described through a Markovian analysis. In order to scale to higher dimension in terms of admissible users and to a variety of radio conditions and to reduce complexity, we propose an approximation closed form model. We assume that contents are divided into chunks of equal small size and that two successive chunks may be retrieved with different radio conditions or from another source, due to mobility. We describe this system by multi-class PS queues.

A rich literature exists on performance models under mobility. Authors in [8] developed Markovian models including

both inter-cell and intra-cell mobilities. However, as the model involves matrix inversion, it is not scalable for systems with a large variety of radio conditions and a large capacity in terms of admissible users. Authors in [9] and [10] introduce closed-form approximations based on Processor sharing and Whittle approximations, respectively. These approximations have good scalability properties, but they do not apply to the extremely flexible case of D2D offloading where users may switch from one link to another dynamically depending on the positions and mobility of caches. Our model copes with these needs by ensuring both scalability and flexibility.

The contributions of this paper can be summarized as follows:

- We develop a scalable product-form approximate performance models for macro and D2D systems, taking into consideration inter- and intra-cell mobility of users. This model fits very well with the exact model in both static and dynamic scenarios.
- We propose a simple offloading strategy exploiting caches embedded within devices where the most popular contents are placed. D2D communications are then exploited for serving neighboring users.
- We evaluate the offloading capacity of the proposed combined content caching and D2D strategy and show that accounting for mobility leads to a better performance.

The rest of the paper is organized as follows. We present the system under consideration in Section II. Then we provide an exact Markovian model of the system for the static case in Section III and for the dynamic case in Section IV. We propose a closed form approximation model in Section V and we present numerical results in Section VI. Finally, we give conclusions in Section VII.

II. SYSTEM DESCRIPTION

We consider a D2D enabled mobile network with moving users. A D2D communication is initiated between two devices if a user requests a content from the Internet and this content is present in the local cache of a neighboring device. Both of them, transmitting and receiving device, are moving. In the case the device with cache goes out from the D2D range, downloading is resumed using D2D if the content is in the cache of another neighboring device, otherwise the downloading is taken in charge by the base station.

A. Spectrum allocation and radio conditions

The spectrum is shared between classical cellular links and D2D links. Let α be the proportion of spectrum allocated to the cellular links; we suppose that the remaining spectrum is shared among D2D links. We also assume that a centralized scheduler allocates the resources to the D2D links within each cell, avoiding thus interference between them.

We decompose the macro cell into J concentric zones, which represent a segmentation of users in J classes, corresponding to different radio conditions, as in Figure 1. We denote by $c_{j,0}$ the standalone rate of users of class j ($j = 1, \dots, J$), corresponding to the transmission rate when

using all the resources of the macro base station, which is $c_{j,0} = f(\text{SINR}_j)$ function of the signal to noise plus interference ratio in zone j . Similarly, we decompose the D2D area into concentric zones considering L classes of D2D links, characterized by the distance between the transmitter and the receiver. In particular, we say that a D2D link is of class l if the transmitter is in zone l . Let c_l , with $l \in [1, \dots, L]$, be the standalone rate of a D2D link of class l when he occupies all the D2D resources.

We denote by Q_1, \dots, Q_J and q_1, \dots, q_L the probabilities for a connection to be of a macro and D2D class (such that $\sum_{j=1}^J Q_j = 1$ and $\sum_{l=1}^L q_l = 1$), respectively, and by R_1, \dots, R_J and r_1, \dots, r_L the macro and D2D zone radii.

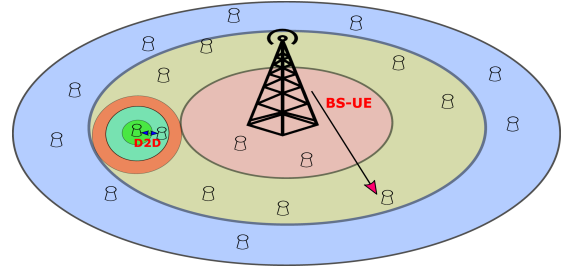


Fig. 1. Decomposition of cell and D2D area.

B. Caching policy

We consider a mobile network where users are endowed by a cache with public contents, known by the base station or a network controller, which they share with other users in the network. For caching solutions, pre-fetching mechanisms are envisaged by pre-staging in caches a bundle of popular contents which are renewed periodically in low traffic periods taking into account popularity changes [3]. We denote by H the hit rate, i.e. the probability to find the content in the cache of a neighboring device.

C. Sojourn/dwell time and downloading time

We consider mobility between devices and mobility with respect to the base station. In particular, for the first type of mobility we can always take the center of the coordinate system at the receiver device and denote by v_T the velocity of the transmitter device with respect to the receiver device. When we consider the mobility of the receiver with respect to the base station we denote by v_R its velocity.

We denote by $t_{k,S}^i$ ($i = d, m$) the sojourn/dwell time for a transmitter in the k -th D2D zone ($i = d$) and for a receiver in the k -th cell zone ($i = m$), respectively. It is distributed exponentially with mean $\mathbb{E}(t_{k,S}^i) = \frac{1}{\eta_k^i}$, see [11]. In particular,

$$\eta_1^d = \eta_1^d \frac{q_1}{q_l} \quad (1)$$

where $\eta_1^d = \frac{2v_T}{\pi r_1}$ with r_1 the radius of the smallest D2D zone, which in our case corresponds to the first D2D circular zone; and

$$\eta_j^m = \eta_1^m \frac{Q_1}{Q_j} \quad (2)$$

where $\eta_1^m = \frac{2v_R}{\pi R_1}$ with R_1 the radius of the first cellular zone. Moreover, we denote by t_D the file downloading time, given by $t_D = \frac{F}{C}$ with F the file size and C the rate at which the file is downloaded.

III. MARKOVIAN MODEL FOR THE STATIC CASE

We start by providing a Markovian description of the system in the static case; mobility is added in the next section. The arrival of new connections is a Poisson process with rate λ uniformly distributed over the cell surface. The file size to be downloaded is exponentially distributed with mean F .

A. D2D communications

We denote by $\vec{x}^d = (x_1^d, \dots, x_L^d)$ the vector in \mathbb{N}^L representing the state of the D2D system, where x_l^d (for $l = 1, \dots, L$) is the number of users of class l .

Transitions between states are caused by the following events in the space of feasible states:

- The arrival of a D2D connection is a Poisson process and, in particular, in ring l it happens with rate $\lambda_l = H\lambda q_l$ and the new state is denoted by \vec{x}_{l+}^d . A new arrival is accepted if $x^d = \sum_{i=1}^L x_i^d \leq N_{max}^d$ with x^d the number of active D2D users in the system and N_{max}^d the maximum number of admissible D2D users.
- When a device finishes downloading the file through a D2D communication in ring l the departure from the system is denoted by \vec{x}_{l-}^d and it depends on the D2D transmission rate $\frac{c_l}{x_l^d}$ in ring l .

Starting from the above assumptions we can construct the generator matrix \mathbf{Q} as follows

$$\begin{aligned} q(\vec{x}^d \rightarrow \vec{x}_{l+}^d) &= H\lambda q_l \\ q(\vec{x}^d \rightarrow \vec{x}_{l-}^d) &= x_l^d \frac{(1-\alpha)c_l}{\sum_{i=1}^L x_i^d F} \end{aligned}$$

where $(1-\alpha)$ is the proportion of resources allocated to D2D communications. Therefore, the steady-state distribution is obtained by solving the following equations: $\vec{\Pi}\mathbf{Q} = 0$ and $\vec{\Pi}\vec{e} = 1$ where $\vec{\Pi}$ is the vector of the steady-state probabilities $\pi(\vec{x}^d)$ and \vec{e} is a vector of ones of appropriate dimension.

B. Macro cell

We denote by $\vec{x}^m = (x_1^m, \dots, x_J^m)$ the vector representing the state of the macro system, where x_j^m represents the number of cellular users of class j .

The events that characterize the evolution of the systems are the following:

- The arrival of a new cellular connection is denoted by \vec{x}_{j+}^m and depends on the probability $(1-H)$ to not find the content in the D2D range. A new arrival is accepted if $x^m = \sum_{i=1}^J x_i^m \leq N_{max}^m$ with x^m the number of active cellular users and N_{max}^m the maximum number of admissible cellular users in the system.

- The end of downloading file in ring j is denoted by \vec{x}_{j-}^m , which corresponds to a departure from ring j , and depends on the macro cell transmission rate $\frac{c_{j,0}}{x_j^m}$.

Therefore, the transition rates for the cellular system corresponding to the above assumptions are:

$$\begin{aligned} q(\vec{x}^m \rightarrow \vec{x}_{j+}^m) &= (1-H)\lambda Q_j \\ q(\vec{x}^m \rightarrow \vec{x}_{j-}^m) &= x_j^m \frac{\alpha c_{j,0}}{\sum_{i=1}^J x_i^m F} \end{aligned}$$

where α is the proportion of spectrum allocated to macro cell.

C. Processor Sharing

The system described above corresponds thus to coupled Processor Sharing queues.

- The D2D system is modeled as a PS queue serving a traffic of $H\lambda F$ and with an overall capacity of $C_D(1-\alpha)$, where

$$C_D = \left[\frac{1}{\lambda} \sum_{l=1}^L \frac{\lambda_l}{c_l} \right]^{-1}$$

is the harmonic average of physical rates for D2D links, supposing that they are allocated all the spectral resources. Its load is thus equal to $\rho_D = \frac{H\lambda F}{C_D(1-\alpha)}$.

- The macro base station is a PS queue serving a traffic of $(1-H)\lambda F$ and with an overall capacity of $C_M\alpha$ (as it used a partial amount of spectral resources). Its load is thus equal to $\rho_M = \frac{(1-H)\lambda F}{\alpha C_M}$ where

$$C_M = \left[\frac{1}{\lambda} \sum_{j=1}^J \frac{\lambda_j}{c_{j,0}} \right]^{-1}$$

is the harmonic average of physical rates for cellular links.

Moreover, we observe that the optimal resource allocation policy is found maximizing the stability condition of the system, given by

$$S(\alpha) = \min \left\{ C_M \frac{\alpha}{1-H}, C_D \frac{1-\alpha}{H} \right\}.$$

Therefore, the optimal value of α is $\alpha^* = \frac{\frac{C_D}{H}}{\frac{C_D}{H} + \frac{C_M}{1-H}}$.

D. Performance evaluation

In order to evaluate the system performance we define the D2D user throughput as follows

$$T_d = \frac{\sum_{\vec{x}^d} \sum_{l=1}^L x_l^d \frac{(1-\alpha)c_l}{\sum_{i=1}^L x_i^d} \pi(\vec{x}^d)}{\sum_{\vec{x}^d} \sum_{l=1}^L x_l^d \pi(\vec{x}^d)}, \quad (3)$$

and the macro user throughput as follows

$$T_m = \frac{\sum_{\vec{x}^m} \sum_{j=1}^J x_j^m \frac{\alpha c_{j,0}}{\sum_{i=1}^J x_i^m} \pi(\vec{x}^m)}{\sum_{\vec{x}^m} \sum_{j=1}^J x_j^m \pi(\vec{x}^m)}. \quad (4)$$

Moreover, the overall user throughput in the system is given by

$$T = (1 - H)T_m + HT_d. \quad (5)$$

IV. MARKOVIAN MODEL WITH MOBILITY

We now consider the case where devices are moving in the system.

A. D2D communications

We consider mobility between devices and, in particular, we assume that the transmitter is moving with velocity v_T with respect to the receiver. The evolution of the system is characterized by the following events:

- Arrival of a new D2D connection in ring l happens again with rate $\lambda_l = H\lambda_{q_l}$. Moreover, with respect to the static case an additional rate, equal to $Hq_l\Lambda_H$, is considered due to the mobility, in particular to users coming from neighboring cell requesting content. The parameter Λ_H is the macro handover rate specified in the following.
- The departure due to the end of downloading is denoted again by \vec{x}_{l-}^d and depends on the D2D transmission rate.
- Migration from ring l to ring $l-1$ ($2 \leq l \leq L$) is denoted by $\vec{x}_{l,l-1}^d$ and happens with rate $\lambda_{l,l-1} = x_l^d \eta_l^d \frac{r_{l-1}}{r_l + r_{l-1}}$ where $\frac{r_{l-1}}{r_l + r_{l-1}}$ represents the probability to move from zone l to $l-1$.
- Migration from ring l to ring $l+1$ ($1 \leq l \leq L$) is denoted by $\vec{x}_{l,l+1}^d$ and it happens with rate $\lambda_{l,l+1} = x_l^d \eta_l^d \frac{r_l}{r_l + r_{l+1}}$.
- Migration of a D2D connection from the border ring to ring l is denoted by $\vec{x}_{L,l}^d$ and depends on the hit rate H .
- Departure of a D2D connection from the border ring L is denoted by \vec{x}_{L-}^d and the rate depends on $(1-H)$, the probability of not finding the content among neighboring devices, and it is given in (6).

The last two transitions mean that when a transmitter goes outside the receiver range, if the content is found in the cache of another neighboring device, the transition happens from the border ring to another ring inside the D2D range, in this way another transmitter takes in charge the transmission. Otherwise, if the content is not found then the receiver will join the cellular communication (contributing to the cellular arrival rate).

We can construct the transition matrix \mathbf{Q} considering the above assumptions:

$$\begin{aligned} q(\vec{x}^d \rightarrow \vec{x}_{l+}^d) &= Hq_l(\lambda + \Lambda_H) \\ q(\vec{x}^d \rightarrow \vec{x}_{l-}^d) &= x_l^d \frac{(1-\alpha)c_l}{\sum_{i=1}^L x_i^d F} \\ q(\vec{x}^d \rightarrow \vec{x}_{l,l-1}^d) &= x_l^d \eta_l^d \frac{r_{l-1}}{r_l + r_{l-1}} \\ q(\vec{x}^d \rightarrow \vec{x}_{l,l+1}^d) &= x_l^d \eta_l^d \frac{r_l}{r_l + r_{l+1}} \\ q(\vec{x}^d \rightarrow \vec{x}_{L,l}^d) &= x_L^d \eta_L^d \frac{r_L}{r_L + r_{L-1}} Hq_l \\ q(\vec{x}^d \rightarrow \vec{x}_{L-}^d) &= \Lambda_{HD} \end{aligned}$$

where Λ_{HD} represents the D2D handover rate and it is given by

$$\Lambda_{HD} = \sum_{\vec{x}^d} x_L^d \eta_L^d \frac{r_L}{r_L + r_{L-1}} (1-H) \pi(\vec{x}^d). \quad (6)$$

The vector of steady-state probabilities $\vec{\Pi}^d = (\pi(\vec{x}^d))$ is obtained by solving the equation:

$$\vec{\Pi}^d \mathbf{Q} = 0. \quad (7)$$

We observe that the handover rate and the steady state probabilities are inter-dependent. This leads to a fixed point calculation of Λ_{HD} .

B. Macro cell

We consider that users are moving in the system with respect to the base station with velocity v_R . The events that characterize the evolution of the systems are the following:

- Arrival of a new cellular connection depends on the probability $(1-H)$ of not finding the content in D2D range and on the D2D handover rate Λ_{HD} . Moreover, due to inter-cell mobility, the arrival rate in the border macro ring J depends also on the handover rate Λ_H . The homogeneity of users in the cell allows us to consider the arrival rate from a neighboring cell equal to the handover rate towards this cell.
- The departure due to the end of downloading is denoted by \vec{x}_{j-}^m and depends on the cellular transmission rate in ring j .
- Migration from ring j to ring $j-1$ is denoted by $\vec{x}_{j,j-1}^m$ ($2 \leq j \leq J$) and happens with rate $\lambda_{j,j-1} = x_j^m \eta_j^m \frac{R_{j-1}}{R_j + R_{j-1}}$, with $\frac{R_{j-1}}{R_j + R_{j-1}}$ the probability to move from zone j to $j-1$.
- Migration from ring j to ring $j+1$ is denoted by $\vec{x}_{j,j+1}^m$ ($1 \leq j \leq J$) and happens with rate $\lambda_{j,j+1} = N_j^m \eta_j^m \frac{R_j}{R_j + R_{j+1}}$.
- Departure of a connection from the border ring J happens with rate Λ_H .

Therefore, the transition rates corresponding to the above assumptions are:

$$\begin{aligned}
q(\vec{x}^m \rightarrow \vec{x}_{j+}^m) &= (1-H)Q_j\lambda + \Lambda_{HD}Q_j \\
q(\vec{x}^m \rightarrow \vec{x}_{j+}^m) &= (1-H)Q_J\lambda + \Lambda_{HD}Q_J + \Lambda_H(1-H) \\
q(\vec{x}^m \rightarrow \vec{x}_{j-}^m) &= x_j^m \frac{\alpha c_{j,0}}{J} \\
&\quad \sum_{i=1}^J x_i^m F \\
q(\vec{x}^m \rightarrow \vec{x}_{j,j-1}^m) &= x_j^m \eta_j^m \frac{R_{j-1}}{R_j + R_{j-1}} \\
q(\vec{x}^m \rightarrow \vec{x}_{j,j+1}^m) &= x_j^m \eta_j^m \frac{R_j}{R_j + R_{j-1}} \\
q(\vec{x}^m \rightarrow \vec{x}_{j-}^m) &= \Lambda_H
\end{aligned}$$

where the handover rate Λ_H depends on the steady-state probabilities and it is given by

$$\Lambda_H = \sum_{\vec{x}^m} x_J^m \eta_J^m \frac{R_J}{R_J + R_{J-1}} \pi(\vec{x}^m). \quad (8)$$

Numerical resolution will give us the steady-state probabilities $\pi(\vec{x}^m)$ in the case of mobility with respect to the base station.

V. APPROXIMATE CLOSED FORM MODEL

Solving the steady-state probability equations (for example equation (7)) can be too cumbersome for a large number of radio conditions and a large number of users. In order to derive a simple approximation, we suppose that contents are divided into chunks of equal small size that are stored in caches and retrieved one after the other. Since devices are moving, for the sake of tractability, we assume that the average radio link quality does not change during the transmission of one chunk, but that two successive chunks may be retrieved with different radio link conditions or from different sources. The original average file size is equal to F and the average chunk size is denoted by F_c . In order to obtain the loads of each queue, we suppose that the content file is of an exponential size so that the number of chunks follows a geometric law of parameter $g = 1 - \exp(-\frac{F_c}{F})$ representing the probability of downloading the whole video. We denote by $\epsilon_n = (1-g)^{(n-1)}$ the probability of needing more than $n-1$ chunks for completing the download.

For the downloading process we consider that chunks can be retrieved from different devices or from the macro base station. Let $\zeta_{k,l}^n$ be the probability the chunk numbered $n+1$ is retrieved using a link of class l knowing that the chunk number n was retrieved using a link of class k ($\sum_l \zeta_{k,l}^n = 1, \forall(n, k)$). $l=0$ corresponds to the macro base station while $l>0$ corresponds to the type of the D2D link characterizing the distance between the user requesting the chunk and the device having it. $n=0$ corresponds to the exogenous arrival of the first chunk, so that $\zeta_{0,l}^0$ corresponds to the probability that the first chunk is served by the macro base station (if $l=0$) or by a D2D link of type $l>0$. Explicit expression of probabilities $\zeta_{k,l}^n$ will be given in the next subsection.

A. Processor sharing modeling

The system corresponds to two coupled Processor Sharing queues, one for the macro base station and the other for the D2D links:

- Macro queue model: Let J be the number of radio conditions in the macro cell, corresponding to the positions of users with respect to the base station. Each class $j \in [1, \dots, J]$ is divided into an infinite number of subclasses, with subclass n corresponding to the transmission of the n -th chunk of the content.
- D2D queue model: Let L be the number of types of D2D links. The D2D queue has $J \times L$ classes, each of them divided into an infinite number of subclasses corresponding to the order of the chunk $n > 0$.

We now compute the traffic load of each subclass numbered (j, l, n) , j standing for the position of the user in the macro cell, l for the serving D2D link and n for the chunk order. Let $\nu_{j,l,n}$ be the arrival rate of chunks in class (j, l, n) . In particular, $\nu_{j,l,n}$ with $l > 0$ corresponds to chunks served by the D2D link, while $\nu_{j,0,n}$ (with $l=0$) corresponds to the ones served by the base station. Therefore, for D2D links ($l > 0$) we have

$$\nu_{j,l,n} = \begin{cases} (\lambda_j + \Lambda_H^{(c)} 1_{j=J}) \zeta_{0,l}^0 \epsilon_1 & \text{if } n = 1 \\ \sum_{k=1}^L \nu_{j,k,n-1} \zeta_{k,l}^{n-1} \epsilon_n & \text{if } n > 1 \end{cases}$$

where $1_{j=J}$ is equal to 1 if $j=J$ and 0 otherwise. As said above, the probabilities $\zeta_{k,l}^{n-1}$ reflect the intra-cell mobility and they will be given in the following. Due to inter-cell mobility chunk requests in macro zone J arrive according to an additional rate $\Lambda_H^{(c)}$ (macro handover rate). It represents the arrival rate of chunk requests of users coming from a neighboring cell. The arrival rate of chunk requests in the D2D zone (j, l) is

$$\nu_{j,l} = \sum_{n=1}^{\infty} \nu_{j,l,n}$$

and the the corresponding load is

$$\rho_{j,l} = \frac{\nu_{j,l} F_c}{(1-\alpha) c_l}.$$

The total load of the D2D queue is:

$$\rho_d = \sum_{j=1}^J \sum_{l=1}^L \rho_{j,l} = \sum_{j=1}^J \sum_{l=1}^L \frac{\nu_{j,l} F_c}{(1-\alpha) c_l} = \sum_{l=1}^L \rho_{l,d} \quad (9)$$

where the last equality is given by rearranging the terms as follows $\rho_{l,d} = \frac{\nu_l F_c}{(1-\alpha) c_l} = \frac{\sum_{j=1}^J \sum_{n=1}^{\infty} \nu_{j,l,n} F_c}{(1-\alpha) c_l}$.

In the case of a cellular link, we have that the arrival rate of a chunk request can be expressed as:

$$\nu_{j,0,n} = \begin{cases} (\lambda_j + \Lambda_H^{(c)} 1_{j=J}) \zeta_{0,0}^0 \epsilon_1 & \text{if } n = 1 \\ \left[\sum_{w=1}^J \nu_{w,0,n-1} \chi_{w,j}^{n-1} + \nu_{j,L,n-1} \zeta_{L,0}^{n-1} \right] \epsilon_n & \text{if } n > 1 \end{cases}$$

with $\chi_{w,j}^{n-1}$ the probability the chunk numbered n will be retrieved using a cellular link of class j knowing that the chunk number $n-1$ was retrieved using a cellular link of class

w. Moreover, the term $\nu_{j,L,n-1}\zeta_{L,0}^{n-1}$ represents the arrival rate of chunk requests in the D2D subclass $(j, L, n - 1)$ which will become requests to base station due to mobility. The term $\Lambda_H^{(c)} 1_{j=J}\zeta_{0,0}^c$ represents chunk requests of users coming from a neighboring cell which are not satisfied by D2D communications. Therefore, the arrival rate of chunk requests in cellular class j is given by

$$\nu_{j,0} = \sum_{n=1}^{\infty} \nu_{j,0,n}.$$

The load of the macro base station is computed by:

$$\rho_m = \sum_{j=1}^J \rho_{j,0} = \sum_{j=1}^J \frac{\nu_{j,0} F_c}{\alpha c_{j,0}}. \quad (10)$$

We observe that we can define the D2D handover rate as

$$\Lambda_{HD}^{(c)} = \sum_{j=1}^J \sum_{n=2}^{\infty} \nu_{j,L,n-1} \zeta_{L,0}^{n-1} \epsilon_n \quad (11)$$

and the macro handover rate as

$$\Lambda_H^{(c)} = \sum_{n=2}^{\infty} \nu_{J,0,n-1} \chi_{J,J+1}^{n-1} \epsilon_n. \quad (12)$$

These handover rates are evaluated iteratively through a fixed point calculation.

The system corresponds to a Kelly network of PS queues with a product form solution for the steady-state probabilities: the stationary distribution of the number of active users in each ring is that of the number of users in a PS queue with classes of respective loads $\rho_{l,d}$ for D2D and $\rho_{j,0}$ for base station:

$$\pi(\vec{x}^d) = \frac{(x_1^d + \dots + x_L^d)!}{x_1^d! \dots x_L^d!} \prod_{l=1}^L (\rho_{l,d})^{x_l^d} (1 - \rho_d), \quad (13)$$

and

$$\pi(\vec{x}^m) = \frac{(x_1^m + \dots + x_J^m)!}{x_1^m! \dots x_J^m!} \prod_{j=1}^J (\rho_{j,0})^{x_j^m} (1 - \rho_m). \quad (14)$$

We give now explicit expressions to the probabilities ζ introduced above. In the case $n = 0$ (which means the first chunk)

$$\zeta_{0l}^0 = \begin{cases} (1 - H) & \text{if } k = 0, l = 0 \\ Hq_l & \text{if } k = 0, l > 0, \end{cases} \quad (15)$$

in the case we consider chunk with order $n > 0$ we have

$$\zeta_{kl}^n = \begin{cases} \gamma_{k,k} & \text{if } k \in]0, L[, l = k \\ \gamma_{k,l} & \text{if } k \in]0, L[, |l - k| = 1 \\ 1 & \text{if } k = l = 0 \\ 0 & \text{if } k = 0, l > 0 \\ \gamma_{L,L} + \psi Hq_L & \text{if } k = L, l = L \\ \gamma_{L,L-1} + \psi Hq_{L-1} & \text{if } k = L, l = L - 1 \\ \psi Hq_l & \text{if } k = L, |l - k| > 1 \\ \psi(1 - H) & \text{if } k = L, l = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\psi = (1 - \gamma(L, L) - \gamma(L, L - 1))$ and $\gamma_{k,l}$ are given in the following. We observe that the last equations represent

that the device from which the content is retrieved may change only when it becomes too far and almost out of coverage, and that the next link may be another D2D link or the macro base station if no other D2D link is available for the same content. Denoting by $t_{k,D}^d$ the chunk downloading time in D2D zone k , given by

$$t_{k,D}^d = \frac{F_c}{(1 - \alpha) \frac{c_k}{x^d}}, \quad (16)$$

we define the probability $\gamma_{k,k}$ to remain in a D2D zone k as the probability that the sojourn time is longer than the chunk downloading time

$$\begin{aligned} \gamma_{k,k} &= \mathbb{P}(t_{k,S}^d > t_{k,D}^d) \\ &= \exp\left\{-\frac{\eta_k^d F_c}{(1 - \alpha) \frac{c_k}{x^d}}\right\} \\ &= \exp\left\{-\frac{\eta_k^d F_c \rho_d}{(1 - \alpha) c_k (1 - \rho_d)}\right\}, \end{aligned} \quad (17)$$

where η_k^d (for $k = 1, \dots, L$) is given in equation (1) and ρ_d is the D2D total load given in equation (9). Moreover, the following relations hold

$$\gamma_{k,k} + \gamma_{k,k+1} + \gamma_{k,k-1} = 1, \quad (18)$$

$$\gamma_{k,k+1} = \gamma_{k,k-1} \frac{2\pi r_k}{2\pi r_{k-1}}. \quad (19)$$

We observe that the equation (18) is related to the fact that a user can remain in a zone or move towards neighboring zones. The equation (19) means that the probability that a D2D link of class k becomes a link of class $k + 1$ is proportional to the perimeter of contact with its two adjacent zones, with radius equal to r_k and r_{k+1} , respectively. The last two conditions above imply

$$\gamma_{k,k-1} = \frac{1 - \gamma_{k,k}}{1 + \frac{r_k}{r_{k-1}}}.$$

In the case without mobility the values of probabilities ζ_{kl}^n are the following:

$$\zeta_{kl}^n = \begin{cases} Hq_l & \text{if } n = 0, k = 0, l > 0 \\ (1 - H) & \text{if } n = 0, k = 0, l = 0 \\ 1 & \text{if } n > 0, k = l \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

Similarly, we define the probabilities χ_{wj}^n describing the mobility with respect to the base station

$$\chi_{wj}^n = \begin{cases} \beta_{w,w} & \text{if } w \in [1, J], j = w \\ \beta_{w,j} & \text{if } k \in [1, J[, |w - j| = 1 \\ \beta_{J,J-1} & \text{if } k = J, l = J - 1 \\ 1 - \beta_{J,J} - \beta_{J,J-1} & \text{if } w = J, j = J + 1 \\ 0 & \text{otherwise} \end{cases}$$

The probability $\beta_{j,j}$ to remain in a macro zone j is defined as the probability that the sojourn time in the zone j is longer than the chunk downloading time in zone j :

$$\begin{aligned}\beta_{j,j} &= \mathbb{P}(t_{j,S}^m > t_{j,D}^m) \\ &= \exp\left\{-\frac{\eta_j^m F_c}{\alpha \frac{c_{j,0}}{x_m^m}}\right\} \\ &= \exp\left\{-\frac{\eta_j^m F_c \rho_m}{\alpha c_{j,0}(1-\rho_m)}\right\}\end{aligned}$$

where η_j^m (for $j = 1, \dots, J$) is given in equation (2) and ρ_m is the macro load given by equation (10). Moreover, the following relations hold

$$\beta_{j,j} + \beta_{j,j+1} + \beta_{j,j-1} = 1, \quad (21)$$

$$\beta_{j,j+1} = \beta_{j,j-1} \frac{2\pi R_j}{2\pi R_{j-1}}. \quad (22)$$

B. Direct computation of the handover rates

We have shown in equation (11) how to compute the handover rate for D2D communications iteratively. We now show that they can be computed directly.

We define the Markov chain describing the evolution of a D2D communication during its lifetime, until it ends its communication or it switches to the macro base station. This Markov chain has $L+2$ states. The states which are numbered from 1 to L are transient and correspond to the L D2D link qualities, the state number 0 is absorbing and corresponds to switching to the macro base station and the state numbered $L+1$ is absorbing and corresponds to the end of the file. In this Markov chain, the transition probabilities between transient states numbered k and l are equal to $p_{kl} = \zeta_{kl}^1(1-g)$, while the transition to the absorbing state $L+1$ occurs with probability g , see Figure 2. The transition to the absorbing state 0 occurs only from state L with probability $\psi(1-H)(1-g)$. The

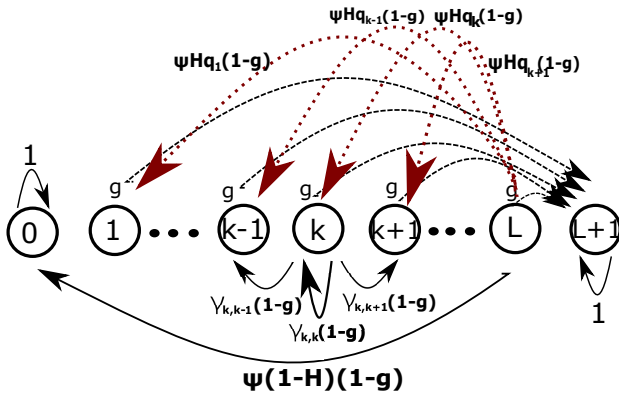


Fig. 2. Markov chain where transition probabilities $p_{k,l}$ are represented with $\psi = (1 - \gamma_{L,L} - \gamma_{L,L-1})$.

probability h_l for a connection that started with a D2D link of quality k switches to the macro base station is equal to the

absorbing probability at state 0. It can be computed by solving the set of equations

$$\begin{cases} h_0 = 1 \\ h_i = (1-g)(\gamma_{i,i-1}h_{i-1} + \gamma_{i,i}h_i + \gamma_{i,i+1}h_{i+1}) & 1 \leq i \leq L-1 \\ h_L = (1-H)\psi(1-g)h_0 + \psi H q_1(1-g)h_1 + \\ \quad + \dots + \psi H q_{L-2}(1-g)h_{L-2} + \\ \quad + (\gamma_{L,L-1} + \psi H q_{L-1})(1-g)h_{L-1} + \\ \quad + (\gamma_{L,L} + \psi H q_L)(1-g)h_L \\ h_{L+1} = 0 \end{cases}$$

Using the hitting probabilities we can compute the D2D handover rate as follows

$$\Lambda_{HD}^{(h)} = \lambda H \sum_{l=1}^L q_l h_l \quad (23)$$

Moreover, the hitting time can be easily computed for the considered system.

VI. NUMERICAL RESULTS

We consider in an urban-outdoor scenario a network with hexagonal cells each of them with radius $R = 500$ m, devices uniformly generated and D2D range $r = 100$ m.

To evaluate transmission standalone rates, we evaluate the SINR (Signal to Noise Ratio) modeling base station pathloss by the ITU model [12]. SINR values are then transformed into capacity values per Resource Block (RB) using link level curves. Simulation parameters are given in Table I. It is worth noting that our evaluation methodology is hybrid, merging simulation for getting SINR and corresponding standalone rates, and queuing theoretical modeling for considering the dynamic behavior of users (arrival of connection demands, mobility). This allows a realistic modeling of the radio interface by simulation and a quick and accurate modeling of the flow level scale. For more information on this methodology, readers may refer to [13].

We divide the hexagonal cell into $J = 2$ concentric zones and the D2D circular area into $L = 2$ concentric zones. The resulting capacities (cellular: $c_{1,0} = 36.3607$ Mbps, $c_{2,0} = 12.4812$ Mbps; D2D: $c_1 = 56.4154$ Mbps, $c_2 = 27.4070$ Mbps) are then evaluated in each zone after computation of the point-per-point SINR, transforming them into throughput values per RB and averaging them over the zone. Moreover, the values of the harmonic mean are given by $C_M = 15.5729$ Mbps and $C_D = 31.4498$ Mbps for macro and D2D, respectively.

We take in the following the hit rate $H = 0.3$ which corresponds to the case of high device density in the cell, according to the values in [3].

A. Validation of the approximation

We start by validating the closed-form approximation by comparing it to the Markovian model. In Figure 3, we present the D2D handover rate while increasing the arrival rate λ , for file size $F = 15$ Mbit and velocity $v_T = 50$ Km/h. In particular, Λ_{HD} is the D2D handover rate computed in

TABLE I
SIMULATION PARAMETERS

Cell Radius	500 m
Macro base station Pathloss	ITU Model
D2D Pathloss	Winner Model
Number of Tx/Rx antennas	MIMO 2×2
Carrier Frequency	2 GHz
Bandwidth	10 MHz

(6) through the Markovian model and the other quantities are computed through the closed-form approximation. $\Lambda_{HD}^{(c)}$ is defined in (11) and computed iteratively and $\Lambda_{HD}^{(h)}$ is computed through the hitting probabilities in (23). The last two expressions depend on the chunk size. We evaluate $\Lambda_{HD}^{(c)}$ for $F_c = 10, 5, 2$ Kbit and we observe that decreasing the chunk size we get a better approximation of Λ_{HD} . Moreover, for $F_c = 2$ Kbit both $\Lambda_{HD}^{(c)}$ and $\Lambda_{HD}^{(h)}$ overlap the curve representing Λ_H confirming the accuracy of the approximation. Figure 4 provides macro handover rate by increasing the

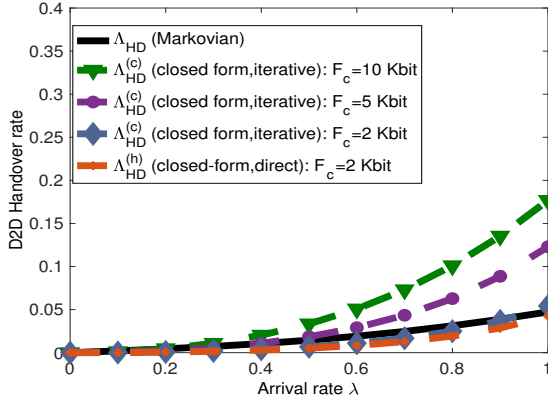


Fig. 3. D2D Handover rate by increasing the arrival rate λ .

arrival rate λ for $v_R = 80$ Km/h, $v_T = 80$ Km/h, file size $F = 15$ Mbit and chunk size $F_c = 4$ Kbit. We observe that the handover rate $\Lambda_H^{(c)}$ computed by the approximation model in (12) fits very well Λ_H evaluated through the Markovian model in (8). While handover rates are essential for the computation of the performance, user throughput is usually the performance metric that is mostly used. Figure 5 illustrates D2D user throughput for different traffic loads. Both results from the exact model and the approximation are illustrated, for a small chunk size ($F = 10$ Mbit and $F_c = 2$ Kbit), as advocated from the handover rates. The proposed approximation is again shown to fit very well the Markovian model.

B. D2D offloading capacity

Having validated the proposed performance evaluation model, we now use it in order to compute the offloading capacity of D2D combined with caching and to assess the impact of mobility on this capacity. We start by illustrating

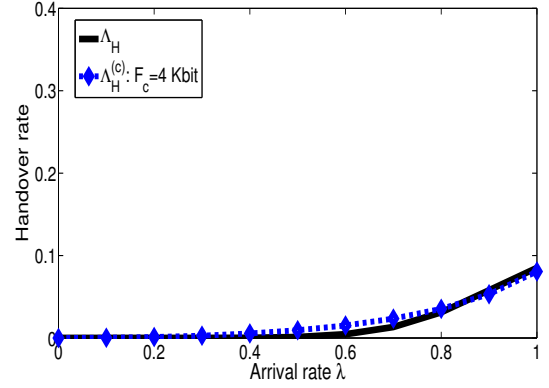


Fig. 4. Macro Handover probability by increasing the arrival rate λ .

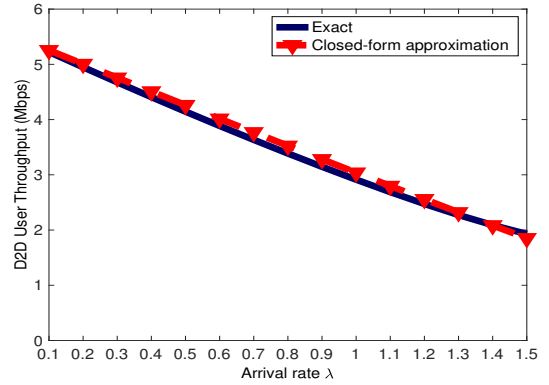


Fig. 5. Validation of the approximation for D2D user throughput.

the impact of mobility on user throughput by comparing the static and dynamic cases. Figure 6 illustrates the D2D user throughput for dynamic and static cases and shows that mobility has a beneficial impact on the throughput of users when served by D2D. Furthermore, Figure 7 plots the overall throughput (averaging over both macro and D2D users) in static and dynamic cases. We also observe that mobility

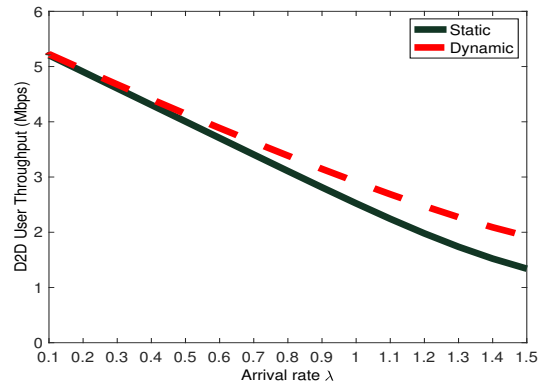


Fig. 6. D2D user throughput in the static and dynamic cases.

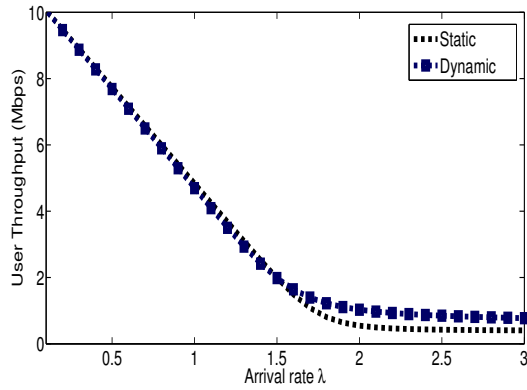


Fig. 7. Global user throughput in the static and dynamic cases.

provides better performance with respect to a static scenario, especially for higher traffic loads. This is due to the fact that users who experience bad radio conditions upon arrival do not accumulate in the system as happens in the static case (see for example the results of [14] for the static case), but have future opportunities for moving to a location with better conditions. Note that this enhancement in performance provided by mobility has been already stated in [15] in the case of wireless ad hoc networks.

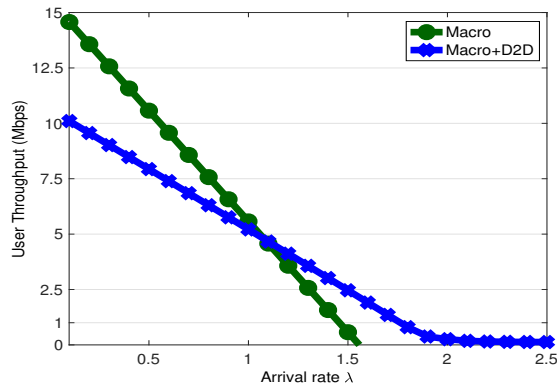


Fig. 8. User throughput in the static case and dynamic cases.

We now move to the evaluation of the offloading capacity of our proposed (D2D+caching) scheme. Figure 8 provides user throughput in the static and dynamic cases. Macro performance (marker-o line) represents the classical system where all users are served via the macro base station. The other curve represents the combination of cellular and D2D communication when devices are endowed with caches. The gain can be observed for the stability condition of the system, representing the largest arrival rate so that the throughput is larger than zero. This maximal capacity is increased from 1.6 to 2 arrivals/second (resulting to a traffic volume increasing from 16 Mbps to 20 Mbps, as each arrival comes with a file of average size 10 Mbit, i.e. a gain of 25%). If the objective of the operator is to ensure a user throughput larger than, say,

1 Mbps, the capacity gain becomes 20%.

VII. CONCLUSIONS

In this work, we have analyzed the offloading capability of D2D communications when devices, endowed by cache, are moving in the network. We look at mobility with respect to the receiving user when we consider D2D communication and with respect to the base station when we consider the classical cellular communication. An exact Markovian model is considered to describe the system. However, in order to scale to higher dimension in terms of number of radio conditions and the number of admissible users, we provide a closed form approximation model based on multi-class PS queues. Numerical results validate the accuracy of our approximation and show that mobility increases the capacity in general and that using D2D communications combined with content caching achieves high capacity gains as it enables offloading traffic to local D2D links.

REFERENCES

- [1] K. Poularakis and L. Tassiulas, "Code, Cache and Deliver on the Move: A Novel Caching Paradigm in Hyper-Dense Small-Cell Networks", IEEE Transactions on Mobile Computing, vol.16, no.3, pp.675-687, March 2017.
- [2] Aymen Jaziri and Ridha Nasri and Tijani Chahed, "Offloading traffic hotspots using moving small cells", IEEE International Conference on Communications (ICC), 2016.
- [3] S.E. Elayoubi, A.M. Masucci, J. Roberts, B. Sayrac, "Optimal D2D Content Delivery for Cellular Network Offloading", ACM/Springer Mobile Networks and Applications (MONET), vol. 22, no. 6, pp. 1033-1044, December 2017.
- [4] R. Wang et al., "Mobility-Aware Caching in D2D Networks", IEEE Transactions on Wireless Communications, vol. 16, no.8, pp. 5001-5015, August 2017.
- [5] R. Wang et al., "Exploiting Mobility in Cache-Assisted D2D Networks: Performance Analysis and Optimization", IEEE Transactions on Wireless Communications, vol.17, no.8, pp. 5592-5605, August 2018.
- [6] Kozyrev D. et al. "Mobility-Centric Analysis of Communication Offloading for Heterogeneous Internet of Things Devices", Wireless Communications and Mobile Computing, vol. 2018.
- [7] M.Chen et al., "Mobility-Aware Caching and Computation Offloading in 5G Ultra-Dense Cellular Networks", Sensors, vol. 16, no. 7, pp. 974-986, June 2016.
- [8] B. Renard, S.E. Elayoubi, A. Simonian, "A dimensioning method for the LTE X2 interface", IEEE Wireless Communications and Networking (WCNC) 2012, Shanghai, China, pp. 2718-2723.
- [9] N. Nya and B. Baynat, "Performance Model for 4G/5G Heterogeneous Networks with Different Class of Users", International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM) 2017, pp. 171-178.
- [10] M. K. Karray, "Users mobility effect on the performance of wireless cellular networks serving elastic traffic", Wireless Networks, vol. 17, no. 1, pp. 247262, 2011.
- [11] B. Jabbari, "Teletraffic aspects of evolving and next-generation wireless communication networks", IEEE Personal Communications 1996, vol.3, no.6, pp.4-9 Dec. 1996.
- [12] ITU-R Report M.2135, "Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced", 2008.
- [13] S.E. Elayoubi, M.K. Karray, Y. Khan, and S. Jeux, S., "A novel hybrid simulation methodology for capacity estimation in mobile networks", Physical Communication, vol. 9, 281-287, December 2013.
- [14] T. Bonald, A. Proutière, "Wireless Downlink Data Channels: User Performance and Cell Dimensioning", Annual International Conference on Mobile Computing and Networking (MobiCom) 2003, San Diego, CA, pp. 339-352.
- [15] M. Grossglauser and D.N.C. Tse, "Mobility increases the capacity of ad hoc wireless networks", IEEE/ACM Transactions on Networking, vol.10, no.4, pp. 477-486, 2002.