Performance of Caching in Device-to-Device Networks with Mobile Nodes

Jarray Chedia, and Chibani Belgacem
University of Gabes, National Engineering School of Gabes, Tunisia
MACS Laboratory : Modeling, Analysis and Control of Systems
jarray.chedia@hotmail.fr, abouahmed17@gmail.com

***Abstract*—A distributed caching device-to-device networks is considered where a user file of interest is stored as several portions in the storage of other devices in the network and obtained successfully through D2D communication. In this paper, we propose that users are mobile and communication links can be spontaneously activated and dropped depending on the users'relative position. By modeling the location of wireless terminal in our network superimposed of independent poisson point process, we derive a service success probability which capture the system performance. Furthermore, the effect of mobility devices, density and file size related to it's type (audio/video) is investigated through numerical results compared to simulation curves**

Keywords—Wireless cache; Device-to-device; Poisson point process; Mobility; File-size.

I. INTRODUCTION

Nowadays,tablets and smatphones have been massively used. The proliferation of these devices has widely increased. They mainly offer confortable and useful viewing experiences. So, data traffic management has becoming more conditioned and very important in our networks. In wireless network, movie streaming currently accounts 50 percent of mobile data traffic during evening hours and is expected to further growth over the next 10 years [2].

In fact, there are the most vision and requirement for 5G, where mobile operators aim to satisfy. For that, major of researchers need to think how they can redesign their current network and seek a new systems concept to boost network capacity, enhance coverage and energy efficiency. For that, one can suffest making content accessible users closer as one solution. So, small base station (BS), that achieve localized communication resources, must deployed [5].

This is frequently affected by social servers, like youtube where different video contents could be seen, at any time and any where, on recent large-screen devices. Nevertheless, many customers are requesting the same video content, which will represent almost the majority of the entire data traffic [8]. The key point of this content reuse problem is the caching strategy, where those popular files are stored in both at mobile device or access point. Obviously, an important consequence is to relieve the overloaded network traffic.

In particular, devices can now cache files that could be shared with the closer users [7]. We consider for this case device-to-device (D2D)communication, in which user's devices located in proximity communicate directly without access to backhaul network [6] (in which transmitter-receiver pairs coexisting in close proximity to establish direct peer -to-peer connections without the use of BS).

All these mechanisms could in fact represent an attractive solution for the encountered challenges arising in the future 5G networks [9].

The idea of caching in the IP networks added to D2D communication paradigm has several benefits compared to traditional architecture [2]-[6]-[3]-[9]-[1]: Firstly, storing content closer to users can offload traffic from cellular network including: radio access network, core network and backhaul. Secondly, the spectral efficiency is improved due to short range between proximate devices, which is the main characteristic of D2D communication. Third, D2D network can be visualized as an ad-hoc network. Finally, the power consumption of UE is reduced with low power links. Therefore, network congestion is simplified, latency is reduced and users's quality-of-experience (QoE)is maximized.

Most of the existing studies assumed a fixed content library, a popularity distribution a priori known and follows the Zipf distribution and a common size of all files. To the best of our knowledge, [4] was the first work to study the problem of optimization placement content, which guarantees maximal total hit probability for multi-coverage regions. However, this work didn't address an optimization of a D2D scenario. In our work, we will revisit the same problem and we will derive a service probability as a function of new parameters such time connection and file's size. Our goal is to find the optimal content placement by deriving the service probability which balances the file size and the connection time for a given file popularity.

The contribution of this article is two fold:

- We first introduce the mobility concept in system model which characterized a D2D communication, but it's more complexity to follow the user's trajectory, therefore , we can simplify this model if we assume that each D2D communication was feasible for a period of time which is known a priori.
- Unlike [2] ,this paper investigates the real case, where the size of memories caches is different from one to other . the storage size depends on the file's size installed. more preciously the file's type (audio or video). It's evident, that the file with interested size, necessitate a lot of time to downloaded. therefore, we characterize the relationship between time connection

and file' size as a function of in terms of service probability of a requested content.

The remainder of this paper is organized as follows: section II describe the system model, in section III, the total service probability is formalized.Finally, we investigate with simulation how various systems parameters, such as the density of D2D devices, time connection affect the service probability. Then, we state and solve the content placement optimization problem. We conclude the paper in section IV.

II. SYSTEM MODEL

We consider a D2D transmitter spatially distributed in $\mathbb{R}^2$ according to an independent homogeneous PPP $\Phi = \{x_i\}$ of intensity $\lambda > 0$, with x_i the location of the $i_{th} - link$ transmitter. Each transmitter communicate with one or many receivers which are modeled also as a PPP with a different density. In this analysis, we will condition on a typical receiver o=(0,0) located at origin of plane, and denote by $r_i = |x_i|$ the distance from each transmitter i to the origin where the receiver resides. The performance metrics are obtained for a test receiver located at the origin, therefore, we drop the notation for the test receiver's location. According to the Slivnyak-Mecke theorem and the stationarity and isotropy of the PPP, these results should be then valid for any generic receiver randomly located on the 2D plane. For the propagation model, we consider the common path loss $l(r_i) = r_i^{-\alpha}$, where α is the path loss exponent. we assume that all transmitters transmit at same power level equal to P and let H be the fading between receiver o and any transmitter at x_i is assumed to be independent and identically distributed. For this analysis, only Rayleigh fading environment is assumed, hence, the channel gain h_i is assumed to be exponentially distributed with unit mean. Therefore, the received power at the tagged receiver located r-meters far away from its transmitters is given by $Ph_i r^{-\alpha}$. If we suppose there no interference between all devices or it is small and the noise power is assumed constant with value σ^2, the quality of coverage at the origin is described by the SNR_o. $\mathrm{SNR}(r_i)$ is the SNR at the reception when receiver o is connected to transmitter $x_i \in \Phi$ and is defined as:

$$\mathrm{SNR}(r_i) = \frac{Ph_i r_i^{-\alpha}}{\sigma^2} \tag{1}$$

A. content and its size

We assume that the receiver o demands for some content (say video, file) from a set of size F that we call a library. It's denoted $C := \{c_1, c_2 ...., c_F\}$ where an element c_j is an available file with size z_j given in bits. We assume that the file's size is not ordered and is dependent upon the type of content, i.e., video, audio, etc. Another important aspect is the relation of file's size with the transmission rate and time connection for each mobile. In fact, the amount of time it takes receiver to download it file request will vary according to the file's size and Internet connection speed. Audio files are very large , probably much larger than most files you work with (unless you work with video). Therefore, a broadband connection is highly recommended for access to music and required for download video. For reference, here are some example of size range and service rate for various media types (see Table I).

Content Type	Size Range (z)	Service Rate (Kbps)
Audio file	120 Kb - 250 Mb	512
Video file	1 Gb - 128 Gb	8000

TABLE I. THE SIZE, AND TRANSMITTED RATE FOR DIFFERENT CONTENT TYPE

B. content popularity law (case for Zipf Distribution)

Furthermore, each content is related to its popularity, which we assume known a priori. The file popularity law describe how a particular content is referenced in an infinite stream of the access requests. Without losing in generality, we assume that this popularity follows a Zipf distribution denoted a_j. Moreover, the probability that the typical receiver will request the content c_j is given by:

$$a_j = \frac{1/j^\gamma}{A} \qquad \text{where} \qquad A = \sum_{j=1}^{F} j^{-\gamma} \tag{2}$$

here A is normalizing constant that gives;

$$\sum_{j=1}^{F} a_j = 1 \tag{3}$$

and γ defines the Zipf exponent which characterize the distribution. The value of γ is depend upon the type of content. Table II, summarize a few reference value for the various content types.

Content Type	Popularity Exponent (γ)
VoD	$0.65 \leq \gamma \leq 1.2$
Websites	$0.64 \leq \gamma \leq 0.83$
P2P Files	$0.75 \leq \gamma \leq 0.82$

TABLE II. SUMMARY OF CONTENT POPULARITY

As discussed in the previous sub-section, each content is characterized by two factors and we can denoted $c_j := (a_j, z_j)$

C. content placement to caches

We consider the case where transmitter have caching memory of size K with a certain pre-installed content. The storage unit of D2D transmitter $x_i \in \Phi$ is denoted by $\Xi_i = \{\xi^1, ..., \xi^K\}$ where an element ξ^j is an entire file from the given library C for the transmitter i. For all transmitter i, the number of elements of Ξ_i should be not larger than K; i.e.,$|\Xi_i| \leq$ K

We consider that the content is independently installed in the caches by same probability distribution which guarantees that

$$b_j = P(c_j \in \Xi_i) \tag{4}$$

i.e the probability that content c_j is stored in any of the memory caches of the network devices is b_j, which are predetermined and known.These probabilities satisfy the following constraints

$$\sum_{j=1}^{F} b_j \leq K \tag{5}$$

$$0 \leq b_j \leq 1, \forall j \tag{6}$$

for the case where the size of memory is expressed in bits denoted by K_b, the sum of size of all files for Ξ_i must not exceed K_b. Then, the constraint becomes

$$\sum_{j=1}^{F} b_j z_j \leq K_b \tag{7}$$

D. User states

In this network system, a experiment was performed allowing the mobility of transmitter devices while the receiver was fixed. Therefore, we assume that all D2D transmitter are allowed to move and availability for communication can change over time, so we will say that each user is in a particular state which can change over time. We can categorized user's state in one of the following types:

- **active user(state 0):** correspond to the typical receiver who request files all the times.

At a given time t, the receiver wants to communicate with another user, then, any transmitter can serve receiver's demand. However, at this time, D2D transmitter is possible for away. As the receiver is assumed to be fixed, he can not follow transmitter's trajectory. To simplify this scenario, we propose that each transmitter has two different state:

- **active transmitter(state 0):** this is a user who is physically close to receiver at time t, and he keeps his position for a duration τ seconds, in which a D2D communication is feasible.

We consider a probabilistic model, where we assign of each transmitter $x_i \in \Phi$ a time of connection denoted τ_i, which are independent and identically distributed according to an exponential distribution.

- **dormant user(state -1):** after time τ_i, this transmitter leaves his position to another farther of typical receiver.

Consequently, the number of active transmitter will vary in time. There are an optimum number of transmitters which can serve receiver o.

III. PERFORMANCE METRICS AND MAIN RESULTS

We assume that the user in the origin request some content from catalog and it is served from the local cache of one or multiple transmitters devices, depending on the availability of the file in the local cache.

For a given content c_j, the receiver is served when at least one transmitter satisfies both conditions *(i)* it has the object c_j and *(ii)* it is sufficiently connected to send this data.

According to what we have explained so far, an instant of the network model within a finite window is given in Figure 1.

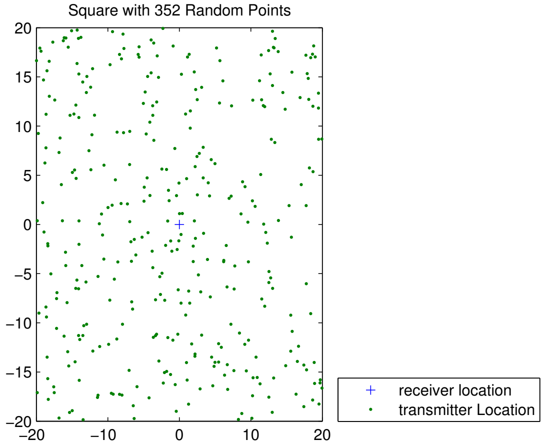


Fig. 1. system model

We define the total service success probability denoted $p_{service}$ as our performance metrics. The total probability of successful service is the expected value of $p_{service,j}$, over the content demand c_j and we write:

$$p_{service} = \sum_{j=1}^{F} a_j p_{service,j} \tag{8}$$

Where a_j is the popularity of content c_j.

The service success probability $p_{service,j}$ refers to the probability that a D2D communication (between typical receiver and at least one transmitter) is successful.

For content c_j chosen by the receiver o with probability a_j, a_j is the popularity of content c_j.
for each transmitter i at point $x_i \in \mathbb{R}^2$, we define:

- $A_{ij} = \{c_j \in \Xi_i\}$: is the event that c_j is installed in his cache inventory and we denote by $p_{ij} = P(A_{ij})$ his probability.
- $B_{ij} = \{\tau_i \ln(1 + \mathrm{SNR}(r)) \geq z_j\}$: is the event that amount of information transmitted from x_i to o within time τ_i (life time of transmission from x_i) is larger or equal than the total size of the data content c_j, equal to z_j and we denote by $q_{ij} = P(B_{ij})$ his probability.

The receiver is served by transmitter i, if

$$\Psi_{ij} := \mathbb{1}_{A_{ij}} \mathbb{1}_{B_{ij}} = 1 \tag{9}$$

i.e. where $\mathbb{1}_{\{...\}}$ is the indicator function which yields 1 if both these independent events occur and 0 otherwise. Consequently, $\forall j \in \{1, ..., F\}$, Service success probability $p_{service,j}$ is defined as

$$p_{service,j} \triangleq p_{ij} q_{ij}, \tag{10}$$

the user o requesting content c_j is served if there is at least one transmitter x_i, who satisfies $\Psi_{ij} = 1$, that is; if the following is true: $\sum_{i=1}^{\infty} \Psi_{ij} \geq 1$ Then the service success probability can be given as follows:

$$p_{service,j} = P[\sum_{i=1}^{\infty} \Psi_{ij} \geq 1] \tag{11}$$

Theorem 1. The probability that a typical receiver can be served by at least one transmitter in the downlink for a certain locally accessible content $j \in \Xi_i$ request can be expressed as:

$$p_{service,j} = 1 - exp(-\pi\lambda b_j \frac{\mathbb{E}[h_i^{2/\alpha}]}{\sigma^{2(2/\alpha)}} \mathbb{E}[(\frac{1}{e^{z_j/\tau_i}-1})^{2/\alpha}]) \quad (12)$$

Remarks:

- From Eq.(12) it is easy to see that the service probability in terms of the desired QoS is a function of: (i) the content placement probabilities b_j, In fact, as b_j increase, $p_{service,j}$ increase. Then, we can control the service probability by varying b_j. In the following we will find the optimal vector $(b_1, ..., b_F)$, that maximize total success probability and avoid the trade-off between most popular files and away for making in caches; (ii) the fraction z_j/τ_i which refers the required rate for content c_j. If the typical receiver request a larger file like very high size's video from Youtyoub (z_j increases), the service probability decreases. However, when a user connects to the typical receiver for a long time (τ_i increases), this last will have more chance to be served and consequently $p_{service,j}$ increases.

- Noting that probability of being served for certain content request depend on the path-loss exponent (α), density of transmitters and their transmit powers. As λ increase, the service probability $p_{service,j}$ approaches 1, This reflects the unrealistic case where all requested files can be cached for each transmitter. Assuming this is not the case in reality due limited size for caches. Consequently, it 's not necessary to be served , the typical receiver must be covered many D2D links.

A. Special cases

The main simplifications we will now consider in various combinations are *(i)* allowing the path loss component $\alpha = 4$, *(ii)* interference is Rayleigh fading with unit mean, which in turn $h_i \sim Exponential(1)$ and *(iii)* the noise power σ^2 is small but non-zero, since in most modern cellular networks thermal noise is not an important consideration, and (iv) file requests of typical receiver are drawn from a library of size J=25 and more precisely from cache memories of size K=1 which can further define by $K_b = z_j$ [Mbits], z_j is the length of file j which will be installed in the cache. This reflects the real case where storage capacity vary according to type of mobile used by the transmitter. For example, memory of Samsung Galaxy S6 (32Go) exceeds that for iPhone 6 (16 Go). Indeed, in this subsection, we consider two specific assumptions corresponding to:

Assumption 1: Time connection is fixed: if we assume that all transmitter $x_i \in \Phi$ connect to typical receiver during T seconds, so $\mathbb{E}[\tau_i] = T$ and Theorem 1 admits a closed form since $\mathbb{E}[(\frac{1}{e^{z_j/\tau_i}})^{2/\alpha}]$ can be evaluated as $(\frac{1}{e^{z_j/T}})^{2/\alpha}$. Therefore, by returning to Eq.(12), service probability can be further simplified to

$$p_{service,j} = 1 - exp[] - \pi\lambda b_j \frac{\Gamma(\frac{2}{\alpha}+1)}{\sigma^{2(2/\alpha)}}(\frac{1}{e^{z_j/T}-1})^{2/\alpha}] \quad (13)$$

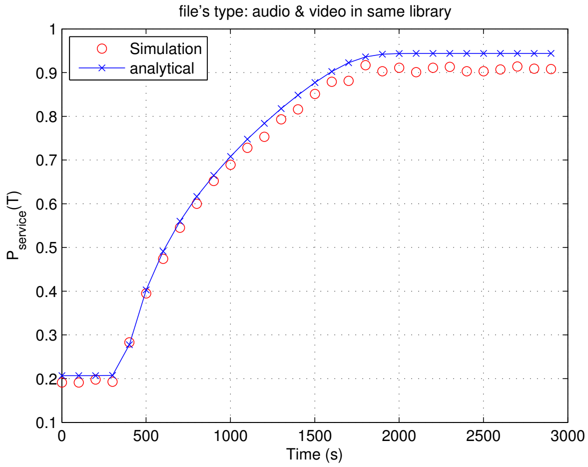


Fig. 2. the total service probability with respect to fixed time connection, for various type of content(video and audio)in same library

Now that we have developed a simple expression for service probability in [13], it is important to validate our approach by showing how these analytical results compare with those of simulation obtained via Monte-Carlo simulations. Figure 2 shows that simulation curves match the theoretical ones.

Assumption 2: Time connection Experiences Exponential Distribution: Significant simplification is possible when the time connection follows an exponential distribution, therefore we can note $\tau_i \sim Exponential(T)$ where T is the rate parameter of the distribution. Referring to table 1, we give the service probability for two case study:

- **case 1:** only audio are stored in the caches, in this case, we assume that receiver o requests content j from the catalog which contains only audio files.
- **case 2:** only video are stored in the caches, all file's size of catalog exceeds 1000 Mbits.
  For those special case,it is possible to write (12) as

$$p_{service,j} = 1 - exp(-\pi\lambda b_j \frac{\Gamma(\frac{2}{\alpha}+1)}{\sigma^{2(2/\alpha)}} I_j) \quad (14)$$

Where

$$\begin{aligned} I_j(T) &= \mathbb{E}[(\frac{1}{e^{z_j/\tau_i}-1})^{2/\alpha}] \\ &= (\frac{1}{e^{z_j/\tau}-1})^{2/\alpha} f_T(\tau) d\tau \\ &= \int_0^\infty (\frac{1}{e^{z_j/\tau}-1})^{2/\alpha} T e^{-T\tau} d\tau \end{aligned} \quad (15)$$

Noting that $f_T(\tau)$ is the probability density of τ exponentially distributed. It is remarkably that expression depends principally on type of contents which was stored (audio or/and file) and time connection to typical receiver, more specifically $\mathbb{E}[\tau]$. Using the expansion $\exp(x) = 1 + x + o(x), x \to 0$ and after an integration of (15), it can be found a closed form of service probability equal to:

$$\begin{aligned} p_{service,j} &= 1 - exp(-\pi\lambda b_j \frac{\Gamma(\frac{2}{\alpha}+1)}{\sigma^{2^{(2/\alpha)}}} I_j) \\ &= 1 - exp(-\pi\lambda b_j \frac{\Gamma(\frac{2}{\alpha}+1)^2}{(z_j T\sigma^2)^{(2/\alpha)}}) \qquad (16) \end{aligned}$$

Where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ is the standard gamma function.

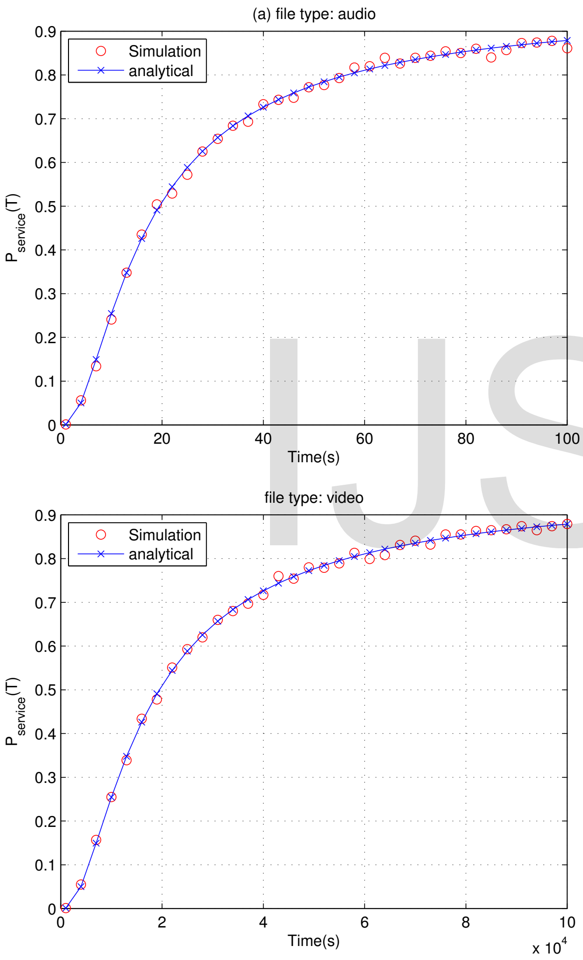


Fig. 3. the total service probability with respect to time connection, for the case where contents installed in caches are (a) audio files and (b) video files.

B. *Validation of the proposed Model*

In the previous section, we have validate our results via numerical simulation and analytical formulas. We observe in fig.2 and 3, that the obtained curves coincide on the majority. On top of these, we are now interested to discuss the impact of several parameters on service probability: (1) density of transmitter λ, (2) connection time τ, and (3) the placement content vector b_j.

2) Impact of transmitter's life time: Time connection, designated also life time in which D2D communication (between receiver o and transmitter at x_i) exists is one critical parameter in our system model. In order to see how it affects the total service provability, plots are given in Figures 2 and 3 respectively. Based on the file type and it corresponding size, each curve represents a different way to build the library. In the figure, we see that service probability increases as we increase the life time. Such a behavior, observed both in theoretical and simulation curves, confirms our intuition in the beginning. The gains become constant after a certain point. This can be explained due to some requested contents are not stored into the caches.

1) Impact of transmitter's density: The evolution of service probability with respect to the number of transmitter's device is depicted in Figure 4. As the transmitter density increases, the receiver o is lucky enough to find at least one transmitter that can serve his request. Therefore, the amount of satisfied requests approaches 1. This increment in service probability can be improved further by increasing the time connection and placement content probability.

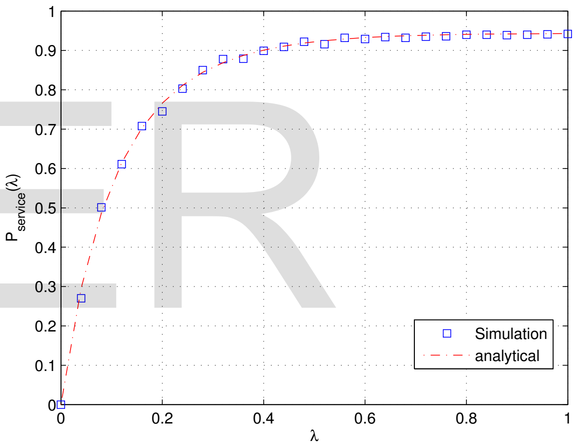


Fig. 4. the total service probability with respect to density λ

3) Impact of placement content: Yet another important parameter in our setup is the content placement probability b_j. Figure 5 shows its impact on service probability for different values of popularity (a_1, a_2), where F=2, time connection is equal to 1000 second, and file size are very large. We can see that we will reach an optimal service probability when we place in the cache memory of size K=1, the most popular content i.e. $b_1 = a_1$ with $a_1 > a_2$. In this part, we focus our intention to $F = 2$ and K=1 for a mathematical simplifications. We can compare the analytical value of b_1^* and that's given by simulation. b_1^* is obtained by solving the following optimization problem:

$$\frac{d}{db_j} p_{service} = 0 \qquad (17)$$

For case F=2, we have:

$b_1 + b_2 = 1$ then $b_2 = 1 - b_1$

$$p_{service}(b_1) = a_1 p_{service,1}(b_1) + a_2 p_{service,2}(1 - b_1) \quad (18)$$

Eq (17). becomes

$$\frac{d}{db_1} p_{service}(b_1) = a_1 \frac{d}{db_1} p_{service,1}(b_1) + a_2 \frac{d}{db_1} p_{service,2}(1 - b_1) \quad (19)$$

leading to:

$$b_1^* = \frac{log_2(\frac{a_2 K_2}{a_1 K_1}) + K_2}{K_1 + K_2} \quad (20)$$

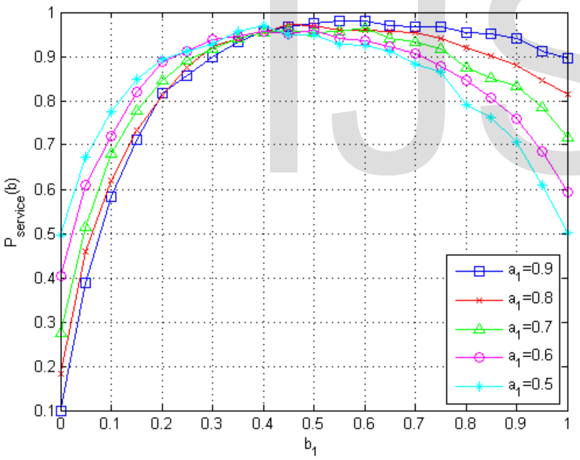
Where,

$$K_1 = -\pi\lambda \frac{\Gamma(\frac{2}{\alpha}+1)}{\sigma^{2(2/\alpha)}} \left(\frac{1}{(e^{z_1/T} - 1)^{2/\alpha}}\right) \quad (21)$$

and,

$$K_2 = -\pi\lambda \frac{\Gamma(\frac{2}{\alpha}+1)}{\sigma^{2(2/\alpha)}} \left(\frac{1}{(e^{z_2/T} - 1)^{2/\alpha}}\right)$$

For the case $F = 2$, we verify that with simulation, given in figure. 6, $p_{service}$ reaches its maximum 0.97 for $b^*_{1,Sim} = 0.45$. This can be calculated thanks to Eq.(20), that shows that, for a given file's popularity $a_1 = 0.8$, an analytical value $b^*_{1,analy} = 0.4681 \approx b^*_{1,Sim}$.



IV. CONCLUSIONS

In this paper, we have shown the importance that has the network paradigm based on the idea of the deployment of caching in devices, to better satisfy the user's request. Indeed, it is through a simple approach based on connection time, we modules the user' mobility. Our results and our analysis provide key information for the deployment of cache-enabled D2D-networks, which can be considered as a promising solution for future ad-hoc networks.

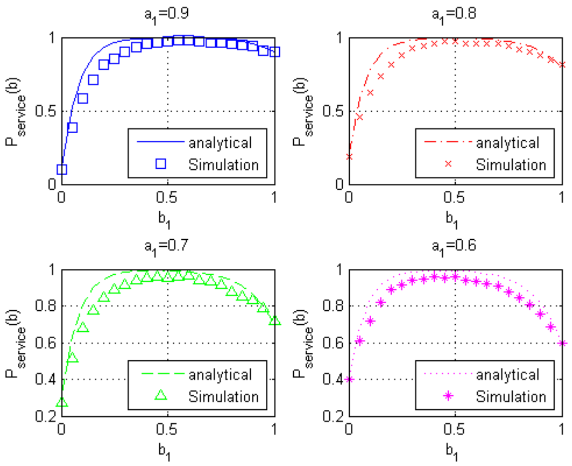


Fig. 6. comparison between the theoretical results and those obtained by simulation

V. ACKNOWLEDGEMENT

This work was supported by the Ministry of the Higher Education and Scientific Research in Tunisia.

REFERENCES

[1] Mehrnaz Afshang, Harpreet S Dhillon, and Peter Han Joo Chong. Modeling and performance analysis of clustered device-to-device networks. *arXiv preprint arXiv:1508.02668*, 2015.

[2] Ejder Bastug, Mehdi Bennis, and Mérouane Debbah. Living on the edge: The role of proactive caching in 5g wireless networks. *Communications Magazine, IEEE*, 52(8):82–89, 2014.

[3] Laszlo A. Belady. A study of replacement algorithms for a virtual-storage computer. *IBM Systems journal*, 5(2):78–101, 1966.

[4] B. Błaszczyszyn and A. Giovanidis. Optimal geographic caching in cellular networks. *IEEE ICC*, 2015.

[5] Federico Boccardi, Robert W Heath, Aurelie Lozano, Thomas L Marzetta, and Petar Popovski. Five disruptive technology directions for 5g. *Communications Magazine, IEEE*, 52(2):74–80, 2014.

[6] Klaus Doppler, Mika Rinne, Carl Wijting, Cássio B Ribeiro, and Klaus Hugl. Device-to-device communication as an underlay to lte-advanced networks. *Communications Magazine, IEEE*, 47(12):42–49, 2009.

[7] Negin Golrezaei, Parisa Mansourifard, Andreas F Molisch, and Alexandros G Dimakis. Base-station assisted device-to-device communications for high-throughput wireless video networks. *Wireless Communications, IEEE Transactions on*, 13(7):3665–3676, 2014.

[8] Negin Golrezaei, Andreas F Molisch, Alexandros G Dimakis, and Giuseppe Caire. Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution. *Communications Magazine, IEEE*, 51(4):142–149, 2013.

[9] Xingqin Lin, Jeffrey Andrews, Amitabha Ghosh, and Rapeepat Ratasuk. An overview of 3gpp device-to-device proximity services. *Communications Magazine, IEEE*, 52(4):40–48, 2014.