Received July 19, 2017, accepted August 14, 2017, date of publication August 21, 2017, date of current version September 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2742555

Mobility-Aware Coded Probabilistic Caching Scheme for MEC-Enabled Small Cell Networks

XINWEI LIU, JIAXIN ZHANG, XING ZHANG, AND WENBO WANG

Wireless Signal Processing and Network Laboratory, Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Xing Zhang (zhangx@ieee.org)

This work was supported in part by the National Science Foundation of China (NSFC) under Grant 61771065, Grant 61571054, and Grant 61631005, and in part by the New Star in Science and Technology of Beijing Municipal Science and Technology Commission (Beijing Nova Program Z151100000315077).

ABSTRACT Caching on the edge has been recognized as an effective solution to tackle the backhaul constraint of network densification. However, most related works ignored user mobility in wireless networks, which is unreasonable under the background of network densification. For a more flexible and context-aware caching decision, the concept of caching on the edge can be extended to mobile edge computing (MEC) that enables computation and storage resources at mobile edge networks. With MEC servers deployed on base stations, a huge amount of collected radio access network context data can be analyzed and utilized to render a caching scheme adaptive to user's context-aware information. In this regard, a novel mobility-aware coded probabilistic caching scheme is proposed for MEC-enabled small cell networks (SCNs). Different from previous mobility-aware caching schemes, user mobility and distributed storage are incorporated into a conventional probabilistic caching scheme, with the aim of throughput maximization. Based on stochastic geometry theory and a modified mobility model of discrete random jumps, the explicit expression of throughput is derived. Due to the complexity of the expression, two light-weight heuristic algorithms are provided to numerically obtain the optimal solutions. Moreover, a significant trade-off among the gains of mobility diversity, content diversity, and channel selection diversity is discussed, and we further numerically analyze how such a trade-off is influenced by user mobility, content popularity, and backhaul capacity, with some fundamental insights into the application of the proposed scheme in MEC-enabled SCNs. The superiority of our proposed scheme is demonstrated by the comparisons with the classical M most popular caching scheme and the conventional probabilistic caching scheme. Numerical results show that our proposed caching scheme achieves higher throughput than those of the other two, especially when users of intense mobility request contents, of which the popularity profile is not skewed, in MEC-enabled SCNs with poor backhaul capacity, indicating that the proposed caching scheme is a promising solution for network densification.

INDEX TERMS Mobility, mobile edge computing, coded caching, probabilistic caching, throughput, stochastic geometry, trade-off.

I. INTRODUCTION

In recent years, with the development and proliferation of mobile devices, we can see an unprecedented global growth of mobile data traffic (reached 7.2 exabytes per month in 2016) [1]. Additionally, it is predicted by [1] that the overall mobile data traffic will grow to 49 exabytes per month by 2021, over three-fourths of which will be video. Network densification is an efficient way to mitigate the aforementioned problem mainly caused by video streaming. With more small base stations (SBSs) installed, pico and femto-cell networks enable a higher capacity of wireless cellular networks [2]. However, the performance of such an approach is under the constraint of the limited capability of backhaul links, which becomes the bottleneck of network densification [3].

Caching on the mobile edge networks has been widely recognized as an economical solution to tackle the aforementioned performance bottleneck of network densification [4], [5]. Equipped with low-cost storage, SBSs can serve mobile subscribers without backhaul congestion if the requested contents are already cached locally, which improves the network performance (e.g., throughput, delay, energy efficiency, etc.) of cache-enabled mobile edge networks [6], [9]. Motivated by this, caching schemes have been widely studied in the scenario of network densification. The authors of [6]–[9] proposed their caching schemes for SBSs, while [10] and [11] studied caching schemes applied to device-to-device (D2D) system. Although these recent works have provided valuable insights into caching decisions in cellular networks and D2D systems, they all assumed that users are fixed nodes in wireless networks, i.e., user mobility is ignored. Obviously, such an assumption is unreasonable under the background of network densification, where SCs are densely deployed and a moving user may thus be served by quite a few SBSs in a short period of time [15].

In order to achieve a higher network performance and better users' quality-of-experience (QoE), the concept of MEC can be utilized to provide a more flexible and context-aware caching decision [12], [13]. Combining computing and caching resources, the MEC servers deployed on BSs are able to analyze and utilize the huge amount of collected RAN context data (e.g., user's location, content popularity and cell sojourn time, etc.) to render a mobility-aware caching scheme [18]. As an intrinsic feature of wireless communication systems, user's mobility pattern has been studied since last century [14], while mobility-aware caching schemes are just studied in recent years. In [15]–[17], optimal storage allocations are proposed considering location transitions of users (e.g., the discrete-time Markov model in [15]) to achieve a higher caching utility [16], throughput [17] and the probability of successful file delivery [15]. However, the intensity of user mobility (e.g., velocity and cell sojourn time) can not be well captured through these location-based models that focus on the user trajectory assumed to be known a prior. Different from location-based models, the authors of [18] and [19] proposed their caching schemes on the basis of inter-contact models that characterize mobility by the length of contact time, which makes it tractable to analyze the impact of mobility intensity on the performance of mobility-aware caching schemes. Generally speaking, analyzing and exploiting the mobility pattern of users, mobility-aware caching schemes achieve higher performance than those of conventional caching schemes (e.g., MPC scheme) [16], [20].

As a data transmission may just continue in a short time period of some minutes in ultra-dense cellular networks [15], a user may possibly download a small portion of the requested file while passing by a BS [20]. Motivated by this, mobility-aware coded caching schemes (i.e., segments of encoded versions of the original file are cached in a distributed manner) are introduced in [15] and [20]. Furthermore, it is numerically demonstrated by [20] that coded caching schemes outperform the uncoded ones. Although [20] has listed some potential factors that affect the performance of a mobility-aware coded caching scheme, e.g., transmission rate, user's sojourn time and the proportion of a coded file cached at BSs, there is a lack of numerical results to analyze how these and other factors influence a mobility-aware coded caching scheme and its performance. Moreover, little work has been done on the mobility-aware coded caching scheme, with the aim of throughput optimization. To the best of our knowledge, this paper is the first work to propose a mobility-aware coded probabilistic caching scheme, aiming to maximize the throughput of ultra-dense cellular networks.

In this paper, we consider a one-tiered SCN equipped with MEC servers for video delivery, where user's mobility feature and content popularity distribution are exploited to render a more context-aware caching decision than that of [7]–[9]. Taking full account of user mobility, content diversity [7] and channel selection diversity [8], we propose a novel mobility-aware coded probabilistic caching scheme with the aim of throughput maximization. Based on the discrete random jump model introduced in [21], we modify the mobility model in terms of sojourn time modeling and define an additional novel model for throughput derivation. Although the explicit expression of throughput is derived, it is hard to further obtain the closed-form optimal solutions due to the complexity of the expression. Hence, two light-weight heuristic algorithms are provided to numerically obtain the optimal solutions. Furthermore, a significant trade-off is derived and we analyze how user mobility, content popularity and backhaul capacity affect the trade-off. Finally, the superiority of our scheme is shown through the comparisons with MPC scheme [9] and probabilistic caching scheme [8]. The main contributions of this paper are summarized as follows:

- We propose a novel mobility-aware coded probabilistic caching scheme for MEC-enabled SCNs, with the aim of throughput maximization.
- For the derivation of throughput, we define a novel model that captures the fundamental coupling between the intensity of mobility and the amount of remaining data of a requested file.
- The explicit expression of throughput is derived and two light-weight heuristic algorithms are provided to numerically obtain the optimal solutions.
- We numerically analyze how user mobility, file popularity and backhaul capacity affect the parameters of our proposed caching scheme, which accounts for the allocation of the gains of mobility diversity, content diversity and channel selection diversity. Comparisons among MPC scheme, probabilistic caching scheme, and the proposed scheme are made to demonstrate that our proposed scheme can be a promising solution to address the challenges of network densification, followed by some fundamental insights into the application of our proposed caching scheme in MEC-enabled SCNs.

The remainder of this paper is organized as follows. Section II describes the system model, including network model and mobility model. In Section III, we introduce the proposed mobility-aware coded probabilistic caching scheme and derive the explicit expression of throughput. In Section IV, the investigated problem is formulated, with the numerical solutions obtained through two light-weight heuristic algorithms. In Section V, numerical results are illustrated and discussed from various aspects, with performance comparisons among three kinds of caching schemes and some fundamental insights. Finally, conclusions are summarized in Section VI.

II. SYSTEM MODEL

In this section, we first introduce the network model, including deployment model, file popularity model, and channel model. Then the mobility model consisting of two parts is introduced, one of which is to characterize user mobility and the other is a novel model for throughput derivation.

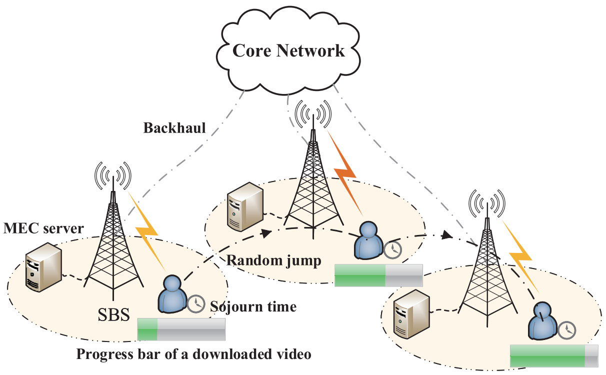

FIGURE 1. An illustration of a one-tiered SCN, where stochastically deployed SBSs are equipped with MEC servers.

A. NETWORK MODEL

1) DEPLOYMENT MODEL

As most of the mobile data traffic is generated by Video-on-Demand (VoD) services [1], we consider a one-tiered SCN for video delivery, where each SBS is equipped with a MEC server of which the storage capacity is C. As shown in Fig. 1, SBSs with low-bandwidth backhaul links are densely and stochastically deployed. Based on stochastic geometry theory, SBSs are modeled as independent homogeneous Poisson point process (PPP) denoted by Φ_s, with the corresponding density of λ_s. Without loss of generality, the bandwidth of all the backhaul links are limited to W_0, while the downlink bandwidth W_s is relatively high (i.e., $W_0 < W_s$). We can thus derive the following equation:

$$W_0 = \theta W_s, \tag{1}$$

where $0 < \theta < 1$ indicates the capacity of backhaul links. A small θ means relative poor backhaul capacity, and vice versa. We henceforth call θ as the *backhaul capability coefficient*.

In order to support real-time High-Definition (HD) video streaming services, all the SBSs are working at the same target downlink spectral efficiency, denoted by ρ. Hence, the downlink transmission rate R_s and the guaranteed transmission rate R_0 are:

$$R_s = \rho W_s, \tag{2}$$

$$R_0 = \rho W_0. \tag{3}$$

2) FILE POPULARITY MODEL

The content library consists of F video files, each of which is of the same size L. As it has been shown in [22] that video access pattern in the Internet follows Zipf's law, we also adopt the Zipf distribution to model file popularity. Arranging videos in descending order of popularity, we denote the set of video indices by $\mathcal{F} = \{1, 2, \ldots, F\}$. The popularity of the i-th ranked video is:

$$f_i = \frac{1/i^{\gamma}}{\sum_{j=1}^{F} 1/j^{\gamma}}, \quad \forall i \in \mathcal{F}, \tag{4}$$

where the parameter γ controls the skewness of file popularity.

3) CHANNEL MODEL

We assume that the transmission power of SBSs is kept constant, denoted by P_t. The standard path loss propagation model is used with path loss exponent $\alpha > 2$. According to [8], given the distance between the reference user and SBS x, denoted by r_x, the signal power received from SBS x is:

$$P_x = P_t |h_x|^2 {r_x}^{-\alpha}, \tag{5}$$

where h_x denotes the Rayleigh fading coefficient. In addition, we assume that a frequency reuse strategy is carefully planned among SBSs, as shown in Fig. 1. Hence, the interfering SBSs are far away from the serving SBS and negligible [8].

B. MOBILITY MODEL

As shown in Fig. 1, user mobility is modeled by discrete random jumps [21], with the corresponding intensity characterized by average sojourn time between jumps. In terms of the distribution of sojourn time, it is reasonable to model it by an exponential function [19], [23], [24]. Therefore, the probability density function (PDF) of sojourn time, denoted by $p(t)$, can be written as follows:

$$p(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}, \quad t \geq 0, \tag{6}$$

where τ is the average sojourn time, which accounts for mobility intensity. A small τ indicates intense mobility (i.e., frequent jumps), and vice versa. In reality, the value of τ can be obtained, with the help of MEC server, by analyzing the huge amount of collected Radio Access Network (RAN) context data through machine learning tools.

We assume that users will continue requesting another video after watching the present video and SBSs serve their users in a round-robin fashion, i.e. a SBS will serve another user when the current file transmission is fulfilled. Considering that the amount of downloaded data is limited by transmission rate and sojourn time [20], a mobile user could only receive a portion of the requested video from a SBS between jumps, as shown in Fig. 1 where the progress bar of a downloaded video gradually grows as the user moves. Let l indicates the remaining data of a requested video at the jumping moment. Consider a general scenario where it takes k jumps for a user to download a video, obviously l

will initially be L when a user starts requesting a video, and then decreases to $L - \sum_{i=1}^{j} d_i$ at the j-th ($0 \leq j \leq k$) jumping moment where d_i indicates the amount of data received between the (i-1)-th and the i-th jump. Finally, l will end at 0 at the k-th jumping moment, and we henceforth refer to the aforementioned process (i.e., l decrease from L to 0) as an *entire download period*. As it is hard to analyze such a complicated scenario, we thus provide a novel model to capture the fundamental coupling between mobility intensity τ and the remaining data of a requested video l. The PDF of l is modeled as follows:

$$p(l) = \frac{e^{\delta \frac{\tau}{T_0} \frac{l}{L}}}{\int_0^L e^{\delta \frac{\tau}{T_0} \frac{l}{L}} dl}, \quad 0 \leq l \leq L, \tag{7}$$

where $\delta > 0$ is a constant determined empirically, controlling the sensitivity of $p(l)$ to mobility intensity, and $T_0 = L/R_s$ is a criterion defined to scale mobility intensity from a statistical perspective. The rationality of the definition will be discussed in Section III-B. Based on the criterion, user mobility with average sojourn time $\tau \leq T_0$ is thought to be intense, and vice versa. Especially when $\tau \ll T_0$ (i.e., $\frac{\tau}{T_0} \approx 0$), it is obvious that a user can just only download a small portion of the requested video file between jumps and thus l can be an arbitrary value between 0 and L with equal possibility at each jumping moment, which indicates a uniform distribution of l (i.e., $p(l) = \frac{1}{L}$). The other extreme case is $\tau \gg T_0$ (i.e., $\frac{\tau}{T_0} \gg 1$), where users will keep contact with the same SBSs for a long time (i.e., user mobility can be ignored in this case). Therefore, users can be treated as fixed nodes, which means the remaining data of each coming request always starts at L (i.e., $p(L) \approx 1$). Obviously, (7) can well characterize the two aforementioned extreme cases. Moreover, it captures the fundamental variation trend between the two cases as τ varies, with δ adjusting the variation sensitivity.

III. PROPOSED CACHING SCHEME AND THROUGHPUT ANALYSIS

A. MOBILITY-AWARE CODED PROBABILISTIC CACHING SCHEME

Our proposed scheme is built on the probabilistic caching scheme in [8], integrated with mobility awareness and distributed storage. Ranked in descending order of popularity, M ($0 \leq M \leq F$) top of all F video files, denoted by Ψ_F^M, are coded and cached in a probabilistic and distributed manner over SBSs. Specifically, for each video $i \in \Psi_F^M$, SBSs independently store the same amount of coded data of video i with the same possibility p ($0 \leq p \leq 1$). By appropriately coding (e.g., MDS code), the coded data of a video cached in SBSs is unique, and a requested video file can be successfully recovered whenever the total amount of downloaded coded data is at least the size of that video [25]. Let m ($0 \leq m \leq 1$) indicates the fraction of coded data of video $i \in \Psi_F^M$ to video size L. In other words, mL coded data of video $i \in \Psi_F^M$ is cached in SBSs with possibility of p. Therefore, a specific caching decision is jointly determined by the m, M and p, which are the parameters of our proposed caching scheme. Assuming that the storage of each SBS is fully utilized, the following equation and inequality are thus derived:

$$MmpL = C, \tag{8}$$

$$F_0 \leq M \leq F, \tag{9}$$

where $F_0 = C/L$ is the lower bound of the integer variable M, which is derived by substituting $m = 1$ and $p = 1$ into (8). Arranging videos in descending order of popularity, let $\mathcal{M} = \{1, 2, \ldots, M\}$ indicates the file indices of Ψ_F^M, and we derive the probability of requesting the cached videos:

$$P_{hit} = \sum_{i \in \mathcal{M}} f_i, \tag{10}$$

where f_i is derived in (4).

B. THROUGHPUT ANALYSIS

We denote the SBSs caching coded data of video i as Φ_i. For $i \in \mathcal{M}$, as SBSs are modeled by PPP with density λ_s, the locations of Φ_i also follow PPP distribution, with density $\lambda_i = p\lambda_s$. To achieve a high transmission rate, the user requesting video i will first attempt to download local coded data from the SBS that offers the strongest received power in Φ_i. The received power is:

$$P_r = \max_{x \in \Phi_i} P_x, \tag{11}$$

where the P_x is the signal power received from SBS x derived in (5). As a frequency reuse strategy is carefully applied, the impact of interference is thus negligible. The corresponding downlink spectral efficiency is thus derived as follows:

$$\eta = \log_2 (1 + \frac{P_r}{\sigma^2}), \tag{12}$$

where σ^2 denotes the noise power variance. A user can successfully connect to a SBS only when the downlink spectral efficiency at least meets the target (i.e., $\eta \geq \rho$). According to the *Theorem 1* derived in [8], for $i \in \mathcal{M}$, the probability that a user requesting video i successfully accesses coded data from Φ_i, with target downlink spectral efficiency ρ is:

$$\begin{aligned} P_c &= 1 - e^{-kp\lambda_s (\frac{P_t}{\sigma^2 (2^\rho - 1)})^{2/\alpha}} \\ &= 1 - e^{-kp\lambda_s (\frac{\xi}{2^\rho - 1})^{2/\alpha}}, \end{aligned} \tag{13}$$

where $k = \pi \frac{\Gamma(\frac{2}{\alpha}+1)}{\Gamma(1)}$, and $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ is the gamma function. $\xi = P_t/\sigma^2$ denotes the signal-to-noise ratio (SNR). If the user requesting video $i \in \mathcal{M}$ fails to access any SBSs of Φ_i, then he or she will attempt to download coded data from the rest of Φ_i, denoted by $\Phi_s \backslash \Phi_i$. As the SBSs of $\Phi_s \backslash \Phi_i$ also distribute in a PPP pattern, with the density $\lambda_{\Phi_s \backslash \Phi_i} = (1-p)\lambda_s$, the corresponding coverage probability is thus derived as follows:

$$P_0 = 1 - e^{-k(1-p)\lambda_s (\frac{\eta}{2^\rho - 1})^{2/\alpha}}. \tag{14}$$

On the one hand, if a user requests a video $i \notin \mathcal{M}$, with the corresponding probability denoted by

P_{miss}, then the situation is relatively simple. The user will be served if it can access a SBS of Φ_s. As the requested video is downloaded through backhaul, only guaranteed transmission rate is available. The coverage possibility and the throughput in this case, denoted by P_s and T_{miss} respectively, are derived as follows:

$$P_s = 1 - e^{-k\lambda_s(\frac{\eta}{2^\rho - 1})^{2/\alpha}}, \tag{15}$$

$$\begin{aligned} T_{miss} &= P_{miss} P_s R_0 \\ &= (1 - P_{hit}) P_s R_0. \end{aligned} \tag{16}$$

On the other hand, in the case where a user requests a video $i \in \mathcal{M}$, the possible situations are various and complicated. For simplicity, we assume that a user will encounter a new SBS as it jumps (i.e, the case of returning to the prior SBS is not considered) during each *entire download period*. Three cases are concluded as follows:

Case I: If the user requesting video i has experienced several jumps before, then the amount of remaining data for video i recovery may become relatively small compared with the amount of coded data stored in Φ_i, i.e., $l \leq mL$. The high transmission rate R_s will achieve if the user successfully accesses a SBS of Φ_i while the guaranteed transmission rate R_0 will be available if the user fails to access any SBS of Φ_i but successfully connects to a SBS of $\Phi_s \backslash \Phi_i$. The throughput of Case I is derived as follows:

$$T_{case1} = P_{hit} P(l \leq mL)[P_c R_s + (1 - P_c)P_0 R_0], \tag{17}$$

where $P(l \leq mL)$ denotes the possibility of $l \leq mL$.

Case II: We consider the case where a user requesting video i successfully access a SBS of Φ_i, with $l > mL$. The user will first download the coded data of video i from local disks of MEC servers. However, as only a portion of the coded video i is cached in a SBS of Φ_i, the user will have to access the remaining coded data through backhaul if it has downloaded mL coded data from the present serving SBS before its next jump. We define $t_0 = mL/R_s$ as the critical sojourn time. Obviously, the user with sojourn time $t \leq t_0$ can enjoy the high transmission rate R_s, while the backhaul-limited transmission will be triggered if $t > t_0$. Hence, it is reasonable to define a criterion to scale the intensity of mobility at the individual level by t_0. Specifically, the user with sojourn time $t \leq t_0$ is thought to be of intense mobility, and vice versa. Plugging $m = 1$ into t_0, we obtain the expression of T_0 that is the criterion defined in Section II-B. Different from t_0 used to scale mobility intensity at individual level, T_0 is a relatively relaxed criterion used at statistic level.

As for $t > t_0$, we define the average transmission rate $R_{avg}(t, l)$:

$$R_{avg}(t, l) = \begin{cases} \dfrac{t_0\, R_s + (t - t_0)R_0}{t}, & t_0 < t \leq t_s(l) \\ \dfrac{t_0\, R_s + [t_s(l) - t_0]R_0}{t_s(l)}, & t > t_s(l)\,, \end{cases} \tag{18}$$

where $t_s(l) = t_0 + (l - mL)/R_0$ is the maximum transmission time of a requested video with remaining coded data l. Obviously, as t increases from t_0, $R_{avg}(t, l)$ will decrease from R_s due to the greater weight coefficient of R_0. When t grows to $t_s(l)$ and still increases, $R_{avg}(t, l)$ will accordingly reach at $R_{avg}[t_s(l), l]$ but stop decreasing. This is because the current VoD request has been served and the SBS will serve another user in a round-robin fashion as mentioned in Section II-B. The throughput of Case II is derived as follows:

$$\begin{aligned} T_{case2} = P_{hit} P_c [&P(l > mL)P(t < t_0)R_s \\ &+ \int_{mL}^{L} \int_{t_0}^{\infty} R_{avg}(t, l)p(t)p(l)\, dtdl], \end{aligned} \tag{19}$$

where $P(l > mL)$ and $P(t < t_0)$ denote the possibility of $l > mL$ and $t < t_0$, respectively.

Case III: We consider the case where a user fails to access the SBSs of Φ_i, with $l > mL$. The user can only download coded data of video i through limited backhaul if it successfully accesses a SBS of $\Phi_s \backslash \Phi_i$. The throughput of Case III is derived as follows:

$$T_{case3} = P_{hit} P(l > mL)(1 - P_c)P_0 R_0. \tag{20}$$

Combining the Case I-III and (16), the throughput of a SBS in MEC-enabled SCNs is:

$$T = T_{case1} + T_{case2} + T_{case3} + T_{miss}. \tag{21}$$

According to (10) and (21), if there are more various videos cached in SBSs (i.e., M increases), the sum cache hit probability P_{hit} will increase, which leads to an increase of throughput. Obviously, the P_c derived in (13) is an increasing function of p. This indicates that a higher caching possibility p of Ψ_F^M will indirectly increase the throughput by P_c, according to (17) and (19). Similar to the definitions of [7] and [8], we term the gains that P_{hit} and P_c bring to throughput as the content diversity gain and the channel diversity gain, respectively. In the same vein, according to (18), (19) and the definition of t_0, a higher m means that a higher R_{avg} will be achieved, and the throughput will increase accordingly. In other words, a large m would allow users with diverse mobility intensity (i.e., either weak or intense mobility) enjoying the high transmission rate R_s. We thus term the gain that R_{avg} contributes to throughput as the mobility diversity gain.

However, under the constraint of Eq. (8), increasing one of the three parameters (i.e., *M, p and m*) of our proposed scheme will result in a decrease of at least one of the other two. As *M, p and m* account for content diversity gain, channel diversity gain and mobility diversity gain respectively, there obviously exists a trade-off among these three kinds of gains. The method to optimally allocate the aforementioned gains will be provided in the next Section.

IV. PROBLEM FORMULATION AND HEURISTIC SOLUTION

The explicit expression of throughput has been derived in Section III-B, which is a three-variable function of the parameters (i.e., m, M and p) of our proposed caching scheme, denoted by $T(m, M, p)$. Substitute (8) into (21), the original function is transformed into a bivariate function of m and M, denoted by $T(m, M)$. However, it is still hard to further obtain the closed-form optimal solutions of (21) due to its complexity. Thus, we numerically obtain the optimal solutions of m and M, denoted by (m^*, M^*), through two classical light-weight heuristic algorithms. The throughput optimization problem can be formulated as follows:

$$\begin{aligned} (m^*, M^*) = \arg\max_{(m,M)} \; & T(m, M), \\ s.t.\; & FL > C, \\ & MmpL = C, \\ & M \in \mathbb{Z}, \\ & F_0 \le M \le F, \\ & 0 \le m \le 1, \\ & 0 \le p \le 1, \end{aligned} \tag{22}$$

where $\mathbb{Z}$ denotes the set of integers. The first constraint $FL > C$ is in accordance with the reality that the total size of content library is larger than the storage of a SBS.

The aforementioned problem can be jointly solved by a particle swarm optimization (PSO) algorithm and a discrete particle swarm optimization (DPSO) algorithm, which are presented in Algorithm 1 and 2 respectively. Specifically, we first obtain m^* with fixed M, denoted by (m^*, M), through a PSO algorithm and then obtain (m^*, M^*) based on (m^*, M) through a DPSO algorithm. Substitute (m^*, M^*) into equation (8), we finally numerically obtain the optimal solutions of m, M and p, denoted by (m^*, M^*, p^*). For a normalized result, we define $\beta = M/F$ the proportion of the content library cached in SBSs. Therefore, (m^*, M^*, p^*) is equivalent to (m^*, β^*, p^*), which accounts for the optimal allocation of mobility diversity gain, content diversity gain and channel diversity gain.

Algorithm 1 DPSO Algorithm for Finding (m^*, M^*)

Input: $T(m, M)$
Output: (m^*, M^*)
for $doi = 1, 2, \ldots N$
  $pbest_i \leftarrow$ a random integer $\in [F_0, F]$
  $v_i \leftarrow$ a random number $\in [-v_{max}, v_{max}]$
end for
for $dok = 1, 2, \ldots maximum\_budget$
  for $doi = 1, 2, \ldots N$
    $p_i \leftarrow [pbest_i + v_i] + 1$
    Limit p_i between F_0 and F
    if $T(m^*, pbest_i) < T(m^*, p_i)$ **then**
      $pbest_i \leftarrow p_i$
    end if
    Update the particles' velocities v_i
    Limit v_i between $-v_{max}$ and v_{max}
  end for
  $gbest \leftarrow \arg\max_{pbest_i} T(m^*, gbest_i)$
  $m \leftarrow m^*$ obtained by **Algorithm 2** with **Input:** $T(m, gbest)$
end for
$(m^*, M^*) \leftarrow (m, gbest)$

Algorithm 2 PSO Algorithm for Finding m^* With Fixed M

Input: $T(m, M)$
Output: m^* and $T(m^*, M)$
for $doi = 1, 2, \ldots N$
  $pbest_i \leftarrow$ a random integer $\in [0, 1]$
  $v_i \leftarrow$ a random number $\in [-v_{max}, v_{max}]$
end for
for $dok = 1, 2, \ldots maximum\_budget$
  for $doi = 1, 2, \ldots N$
    $p_i \leftarrow pbest_i + v_i$
    Limit p_i between 0 and 1
    if $T(pbest_i, M) < T(p_i, M)$ **then**
      $pbest_i \leftarrow p_i$
    end if
    Update the particles' velocities v_i
    Limit v_i between $-v_{max}$ and v_{max}
  end for
  $gbest \leftarrow \arg\max_{pbest_i} T(gbest_i, M)$
end for
$m^* \leftarrow gbest$

V. NUMERICAL RESULTS AND DISCUSSION

In this section, we numerically show the trade-off among the gains of content diversity, channel selection diversity as well as mobility diversity, and discuss how user mobility, file popularity and backhaul capability affect the optimal allocation of the three kinds of gains determined by the parameters of the proposed scheme. Furthermore, comparisons among classical MPC scheme, conventional probabilistic caching scheme [8], and our proposed caching scheme are made to demonstrate the superiority of our scheme in addressing the challenges of network densification. The simulation parameter settings are listed in TABLE 1.

TABLE 1. Simulation parameters.

Parameter	Value	Parameter	Value
λ	0.003 (units/m^2)	SNR	45 (dB)
α	4	δ	1
F	300	ρ	0.8 (bits/s/Hz)
L	200 (MB)	C	12000 (MB)
W_s	10 (MHz)	θ	0.2, 0.5(default), 0.8

A. IMPACT OF USER MOBILITY ON PROPOSED CACHING SCHEME

The impact of user mobility on our proposed scheme is illustrated in Fig. 2, where the *backhaul capability coefficient* θ is set to be $\theta = 0.5$, indicating medium capability of backhaul links. As shown in in Fig. 2(a), m^* has a monotonically increasing behavior with the increase of τ, which means SBSs are desirable to cache a larger proportion of coded data of each video file in Ψ_F^M as the intensity of user mobility becomes weaker. Moreover, we note that such an increasing behavior is more sensitive to the increase of τ with higher content popularity skewness γ.

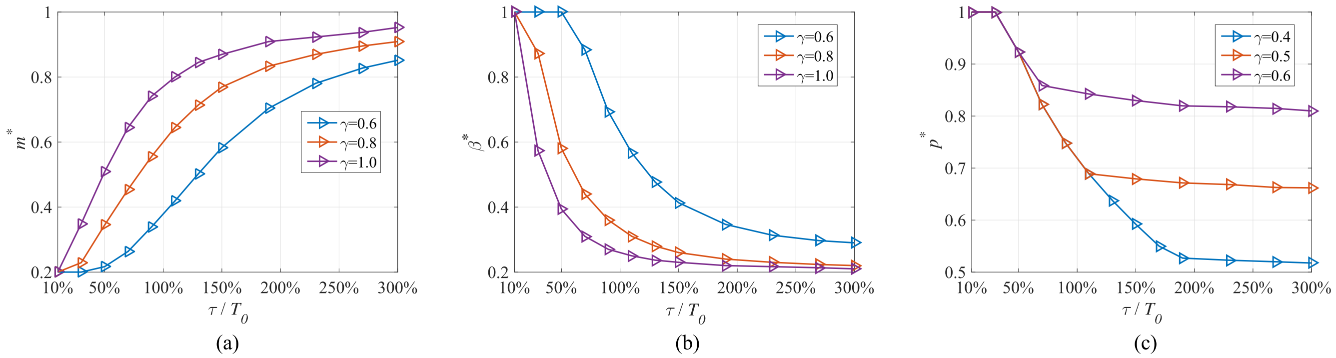

FIGURE 2. **The impact of user mobility on the parameters of proposed caching scheme under different content popularity skewness, with medium backhaul capability ($\theta = 0.5$). (a) m^* vs. the intensity of mobility. (b) β^* vs. the intensity of mobility. (c) p^* vs. the intensity of mobility.**

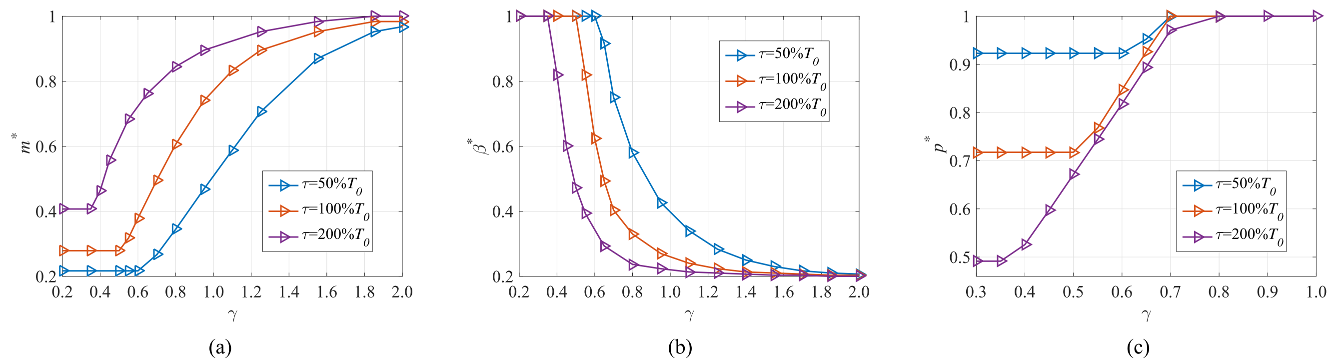

FIGURE 3. **The impact of content popularity on the parameters of proposed caching scheme under different mobility intensity, with medium backhaul capability ($\theta = 0.5$). (a) m^* vs. the content popularity skewness. (b) β^* vs. the content popularity skewness. (c) p^* vs. the content popularity skewness.**

This indicates that precise outcomes of user's mobility pattern are significant when the intensity of mobility is medium (e.g., $50\% \leq \frac{\tau}{T_0} \leq 150\%$) because a small bias in τ will lead to a large deviation in m^*, especially when the popularity profile is skewed. In contrast to m^*, both β^* and p^* decrease monotonically and finally tend to be constant as τ increases, shown in Fig. 2(b) and Fig. 2(c) respectively. However, with an increase in content popularity skewness γ, the decreasing behaviors of β^* and p^* show opposite variation trend. Similar to that of m^*, the decreasing behaviors of β^* becomes more sensitive to the increase of τ as γ becomes larger, while p^* experiences a smaller fluctuation in general with a larger γ. Fundamental conclusions can be summarized as follows:

- It is worth sacrificing the content diversity gain (especially with a skewed popularity profile) and the channel diversity gain (especially with a flat popularity profile) to guarantee the mobility diversity gain as the intensity of user mobility becomes weaker;
- However, blindly sacrificing the two gains is not a wise strategy as the loss will outweigh the gain if β^* and p^* further decrease instead of maintaining at a reasonable level.
- Applying our proposed scheme in MEC-enabled SCNs, precise statistical results of mobility intensity and the accurate prediction of its future value are significant when users in MEC-enabled SCNs are statistically of medium mobility and the popularity profile is skewed.

B. IMPACT OF CONTENT POPULARITY ON PROPOSED CACHING SCHEME

Fig. 3 shows the impact of content popularity on our proposed scheme, where the *backhaul capability coefficient* θ is also set to be $\theta = 0.5$. In Fig. 3(a), m^* shows an upward trend and finally tends to its upper bound with the increase of content popularity skewness γ. Similarly, p^* experiences a relatively sharp increase and finally reaches its upper bound as γ increases, shown in Fig. 3(c). Furthermore, we note that both the increasing behaviors of m^* and p^* become more sensitive to the increase of γ as the intensity of mobility becomes weaker (i.e., with an increase in τ). On the contrary, as illustrated in Fig. 3(b), β^* shows a downward trend and finally tends to its lower bound with the increase of γ. Such a decreasing behavior is also shown to be more sensitive to the increase of content popularity skewness as the intensity of mobility becomes weaker. Fundamental conclusions can be summarized as follows:

- Exchanging content diversity gain for channel selection diversity gain and mobility diversity gain is desirable as the popularity profile becomes more skewed, especially when users are of intense mobility;

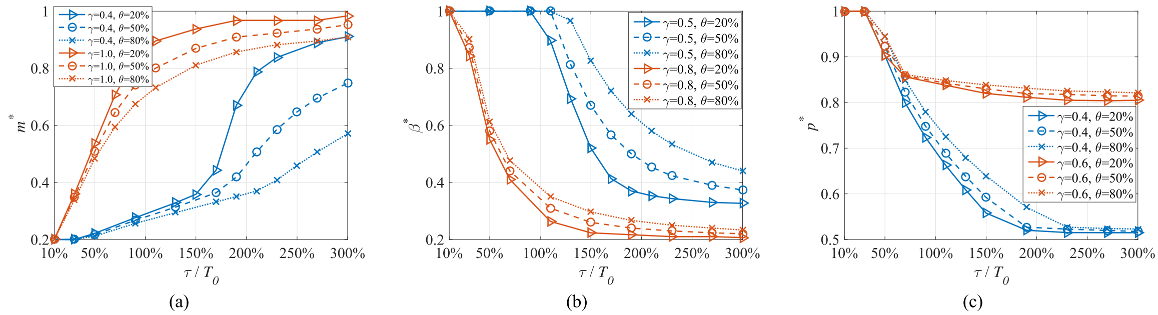

FIGURE 4. **The impact of backhaul capability on the parameters of proposed caching scheme under various mobility intensity and content popularity skewness. (a) m^* vs. the content popularity skewness. (b) β^* vs. the content popularity skewness. (c) p^* vs. the content popularity skewness.**

- However, the loss will outweigh the gain if such an exchange is still conducted with a further increase in the skewness of content popularity. Hence, all the three parameters of the proposed scheme finally tend to be constant.
- Compared with m^*, p^* is much more sensitive to the variation of content popularity skewness, which indicates a higher priority to guarantee channel selection gain rather than mobility diversity gain as the popularity profile becomes more skewed.

C. IMPACT OF BACKHAUL CAPABILITY ON PROPOSED CACHING SCHEME

The impact of backhaul capability on our proposed scheme is plotted in Fig. 4, where the *backhaul capability coefficients* are set to be $\theta = 0.2, 0.5$ and 0.8, corresponding to low, medium and high capability of backhaul links. As illustrated in Fig. 4(a), m^* shows an upward trend with the decrease in the *backhaul capability coefficient* θ, corresponding to the decrease in the capability of backhaul links. In addition, the degree of such an upward trend first gets higher and then gradually decreases as the intensity of mobility becomes weaker, especially with a skewed popularity profile (i.e., large γ). On the contrary, as shown in Fig. 4(b) and Fig. 4(c), both β^* and p^* have decreasing behaviors with the decrease in the capability of backhaul links, the degree of which gets higher and then decreases with an increase in τ. Similar to that of m^*, the two decreasing behavior is more apparently observed with a relatively flat popularity profile. Furthermore, compared with m^* and β^*, p^* is not sensitive to the variation of backhaul capability. For instance, when the content popularity skewness $\gamma = 0.6$, little gap on p^* is observed between low backhaul capability (i.e., $\theta = 0.2$) and high backhaul capability (i.e., $\theta = 0.8$). Fundamental conclusions can be summarized as follows:

- The capability of backhaul links indeed has an important impact on the optimal allocation of the three kinds of aforementioned diversity gains, and it is desirable to sacrifice the content diversity gain and the channel diversity gain to guarantee the mobility diversity gain as the backhaul capability gets worse;

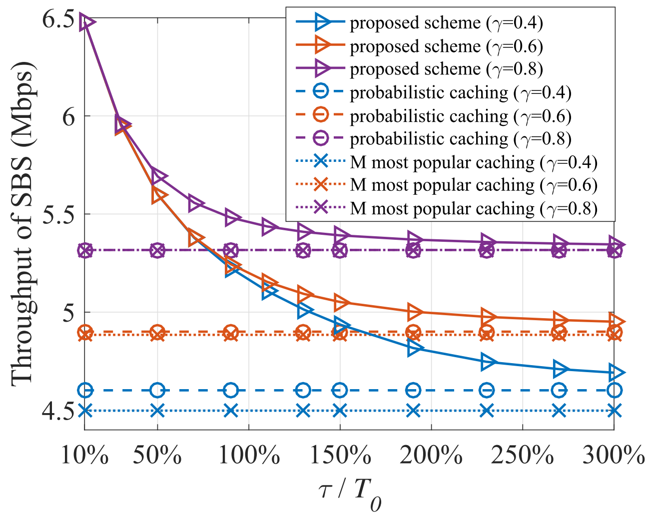

FIGURE 5. **Comparisons among MPC, probabilistic caching and the proposed caching scheme are made under various mobility intensity, content popularity skewness, with medium backhaul capability ($\theta = 0.5$).**

- However, such an impact is weak and even negligible when users are of intense mobility and the popularity profile is skewed.

D. PERFORMANCE EVALUATION

Comparisons among our proposed scheme and the other two widely known caching schemes are plotted in Fig. 5 to evaluate the performance of the proposed caching scheme. Therein, the probabilistic caching scheme in [8] that jointly considers both content diversity and channel selection diversity outperforms MPC scheme, especially with a small content popularity skewness. Compared with probabilistic caching scheme, our proposed scheme further exploits user mobility and distributed storage, outperforming the other two schemes in terms of throughput. Especially when user mobility is intense, our proposed scheme achieves a much higher throughput, which is more remarkable with smaller content popularity skewness. However, we note that the performance of our scheme shows a downward trend and finally tend to that of the probabilistic caching scheme as τ increases.

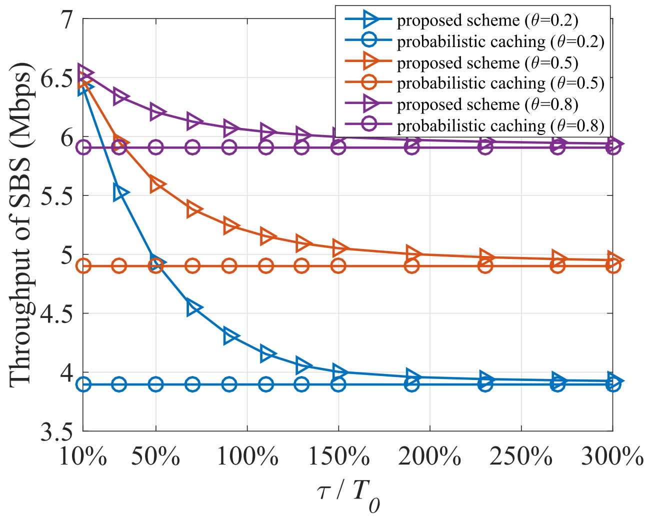

FIGURE 6. **Comparisons between probabilistic caching and the proposed caching scheme are made under various mobility intensity and backhaul capability, with content popularity skewness $\gamma = 0.6$.**

This is due to the diminished advantage of distributed storage with the decrease of mobility intensity. Additionally, a higher γ will accelerate the aforementioned downward trend because the m^* will become larger as γ increase, shown in Fig. 3(a). Obviously, a large m^* will result in a similar caching decision to that of probabilistic caching scheme as distributed storage is not considered in the latter (i.e., $m \equiv 1$). Therefore, the lower bound of the performance of our proposed scheme is that of the conventional probabilistic caching scheme.

In Fig. 6, performance comparisons between the conventional probabilistic caching scheme and our proposed scheme are made with low, medium and high backhaul capability (i.e., $\theta = 0.2, 0.5, 0.8$). Therein, MPC scheme is omitted as probabilistic caching scheme outperforms it, shown in Fig. 5. With a decrease in θ, the performances of the two caching scheme show a downward trend, which is in accordance with the fact that a poor backhaul capability will limit the performance of wireless networks. Furthermore, we note that the performance gap between probabilistic caching scheme and our proposed scheme becomes larger with a decrease in θ and τ, which indicates that our proposed caching scheme is a promising way to mitigate the limited backhaul problem of network densification by exploiting user mobility and distributed storage. For instance, when the average sojourn time $\tau = 25\% T_0$, our proposed caching scheme with low backhaul capability (i.e., $\theta = 0.2$) provides 42.0% and 12.9% higher throughput than those of the probabilistic caching scheme with the low and medium backhaul capability respectively, while the probabilistic caching scheme can only outperform ours by just 6.76% at the cost of a high backhaul capability. Fundamental conclusions can be summarized as follows:

- Our proposed scheme achieves higher throughput than those of probabilistic caching scheme and MPC scheme under various circumstances (i.e., different degree of user mobility, content popularity skewness and backhaul capability), and the performance gap is remarkable with flat popularity profile, intense user mobility and poor backhaul, the last two of which match the characteristics of network densification.
- Applying our proposed scheme in MEC-enabled SCNs, it is unnecessary to increase the throughput by enchanting backhaul capability in the area, where users are mostly of intense mobility due to the diminished impact of backhaul capability on the performance of our proposed scheme under this circumstance.

VI. CONCLUSION

This paper is the first work to propose a mobility-aware coded caching scheme for throughput optimization in dense cellular networks. Built on conventional probabilistic caching, a novel mobility-aware coded probabilistic caching scheme is proposed for video delivery in MEC-enabled SCNs. Different from previous works, content diversity, channel selection diversity and user mobility are jointly considered in our proposed scheme. The explicit expression of throughput is derived based on the modified discrete random jump model. Due to the complexity of the expression of throughput, it is hard to further obtain the closed-form optimal solutions, and two light-weight heuristic algorithms are thus provided for the corresponding numerical solutions. The trade-off among the gains of content diversity, channel selection diversity and mobility diversity is numerically discussed. Furthermore, how user mobility, content popularity and backhaul capability influence such a trade-off is numerically analyzed, followed by some fundamental insights provided for the application of our proposed caching scheme in MEC-enabled SCNs. Compared with typical MPC scheme and conventional probabilistic caching scheme, it is found that our proposed scheme achieves a higher throughput than those of the other two under different degree of user mobility, content popularity skewness and backhaul capability, which is more remarkable with intense user mobility, flat popularity profile and poor backhaul capability. This indicates that our proposed caching scheme is a promising way to address the challenges of network densification.

REFERENCES

[1] Cisco. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021*. [Online]. Available: http://www.cisco.com/go/vni
[2] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 64, no. 9, pp. 59–67, Sep. 2008.
[3] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks: Challenges and research advances," *IEEE Netw.*, vol. 28, no. 6, pp. 6–11, Nov./Dec. 2014.
[4] D. Liu, B. Chen, C. Yang, and A. F. Molisch, "Caching at the wireless edge: Design aspects, challenges, and future directions," *IEEE Commun. Mag.*, vol. 64, no. 9, pp. 22–28, Sep. 2016.
[5] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.
[6] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," in *Proc. IEEE INFOCOM*, Dec. 2012, pp. 1107–1115.

[7] Z. Chen, J. Lee, T. Q. S. Quek, and M. Kountouris, "Cooperative caching and transmission design in cluster-centric small cell networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3401–3415, May 2017.

[8] S. H. Chae and C. Wan, "Caching placement in stochastic wireless caching helper networks: Channel selection diversity via caching," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6626–6637, Oct. 2016.

[9] J. Zhang, X. Zhang, M. A. Imran, B. Evans, and W. Wang, "Energy efficiency analysis of heterogeneous cache-enabled 5G hyper cellular networks," in *Proc. IEEE GLOBECOM*, Dec. 2016, pp. 1–6.

[10] J. Rao, H. Feng, C. Yang, Z. Chen, and B. Xia, "Optimal caching placement for D2D assisted wireless caching networks," in *Proc. IEEE ICC*, May 2016, pp. 1–6.

[11] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless D2D networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 849–869, Feb. 2016.

[12] A. Ahmed and E. Ahmed, "A survey on mobile edge computing," in *Proc. IEEE ISCO*, Jan. 2016, pp. 1–8.

[13] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[14] I. F. Akyildiz, J. McNair, J. S. M. Ho, H. Uzunalioglu, and W. Wang, "Mobility management in next-generation wireless systems," *Proc. IEEE*, vol. 87, no. 8, pp. 1347–1384, Aug. 1999.

[15] K. Poularakis and L. Tassiulas, "Exploiting user mobility for wireless content delivery," in *Proc. IEEE ISIT*, Jul. 2013, pp. 1017–1021.

[16] Y. Guan, Y. Xiao, H. Feng, C.-C. Shen, and L. J. Cimini, "MobiCacher: Mobility-aware content caching in small-cell networks," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 4537–4542.

[17] A. Shabani, S. P. Shariatpanahi, V. Shah-Mansouri, and A. Khonsari, "Mobility increases throughput of wireless device-to-device networks with coded caching," in *Proc. IEEE ICC*, May 2016, pp. 1–6.

[18] C. Min *et al.*, "Mobility-aware caching and computation offloading in 5G ultra-dense cellular networks," *Sensors*, vol. 16, no. 7, p. 974, 2016.

[19] C. Jarray and A. Giovanidis, "The effects of mobility on the hit performance of cached D2D networks," in *Proc. IEEE WiOpt*, May 2016, pp. 1–8.

[20] R. Wang, X. Peng, J. Zhang, and K. B. Letaief, "Mobility-aware caching for content-centric wireless networks: Modeling and methodology," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 77–83, Aug. 2016.

[21] S. Shin, U. Lee, F. Dressler, and H. Yoon, "Analysis of cell sojourn time in heterogeneous networks with small cells," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 788–791, Apr. 2016.

[22] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, S. Moon, *et al.*, "I tube, You Tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. ACM SIGCOMM Conf. Internet Meas.*, 2007, pp. 1–14.

[23] Y. Du, J. Fan, and J. Chen, "Experimental analysis of user mobility pattern in mobile social networks," in *Proc. IEEE WCNC*, Mar. 2011, pp. 1086–1090.

[24] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnovic, "Power law and exponential decay of intercontact times between mobile devices," *IEEE Trans. Mobile Comput.*, vol. 9, no. 10, pp. 1377–1390, Oct. 2010.

[25] D. Leong, A. G. Dimakis, and T. Ho, "Distributed storage allocations," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4733–4752, Jul. 2012.

XINWEI LIU received the B.E. degree in communication engineering from Xidian University, Xi'an, China, in 2016. He is currently pursuing the M.E. degree with the Key Laboratory of Universal Wireless Communications, School of Information and Communication Engineering, Beijing, China. His research interests are mainly in wireless communications and networks, green communication, and mobile edge computing.

JIAXIN ZHANG received the B.E. degree in information engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2012, and the Ph.D. degree from the Key Laboratory of Universal Wireless Communications, School of Information and Communication Engineering, in 2017. His research interests are mainly green communication, mobile edge computing, and 5G network architecture.

XING ZHANG is currently a Full Professor with the School of Information and Communications Engineering, Beijing University of Posts and Telecommunications, China. His research interests are mainly in 5G wireless communications and networks, green communications, cognitive radio and cooperative communications, big data, and Internet of Things. He is the author and coauthor of two technical books and more than 100 papers in top journals and international conferences and filed more than 30 patents. He is a Senior Member of IEEE ComSoc and a member of CCF. He received the Best Paper Award in the 9th International Conference on Communications and Networking in China (2014), the 17th International Symposium on Wireless Personal Multimedia Communications (2014), and the 8th IEEE International Conference on Wireless Communications and Signal Processing (2016). He has served as the General Co-Chair of the 3rd IEEE International Conference on Smart Data (SmartData-2017), as a TPC co-chair and a TPC member for a number of major international conferences, including MobiQuitous 2012, IEEE ICC/GLOBECOM/WCNC, CROWNCOM, and Chinacom.

WENBO WANG received the B.S., M.S., and Ph.D. degrees from the Beijing University of Posts and Telecommunications (BUPT) in 1986, 1989, and 1992, respectively. He is currently a Professor and the Executive Vice Dean of the Graduate School, BUPT. He is currently the Assistant Director of the Key Laboratory of Universal Wireless Communication, Ministry of Education. He has published more than 200 journal and international conference papers, and six books. His current research interests include radio transmission technology, wireless network theory, and software radio technology.