Received September 14, 2016, accepted October 8, 2016, date of publication December 1, 2016, date of current version January 4, 2017.

Digital Object Identifier 10.1109/ACCESS.2016.2633485

Challenges of Mobile Social Device Caching

YECHENG WU[1], SHA YAO[2], YANG YANG[1], TING ZHOU[1], HUA QIAN[2,3], HONGLIN HU[2,3], AND MATTI HAMALAINEN[4]

[1]Key Laboratory of Wireless Sensor Network and Communications, Shanghai Institute of Microsystem and Information technology, Chinese Academy of Sciences, Shanghai 20035, China
[2]Shanghai Research Center for Wireless Communication, Shanghai 200035, China
[3]Shanghai Advanced Research Institute, Chinese Academy of Sciences Shanghai 200035, China
[4]DIMECC, Tampere 733720, Filand

Corresponding author: Y. Yang (yang.yang@wico.sh)

This work was supported in part by the National Natural Science Foundation of China under Grant 61571004, in part by the Technology Commission of Shanghai Municipality under Grant 14ZR1439600, in part by the international cooperative project of the Ministry of Science and Technology under Grant 2014DFE10160, in part by the Ministry of Science and Technology 863 Hi-Tech Program under Grant 2015AA01A702, and in part by the Popular Science Education platform of Online electric experiment for teenagers under Grant 16DZ2348600.

ABSTRACT Wireless caching at network edges (in user devices or in base stations) is considered to be a promising solution to alleviate backhaul overload in future wireless networks. Device-to-Device (D2D)-assisted caching emerges as an attractive option. Due to the nature of D2D communications, social networks and the characteristics within are intertwined with the challenges in mobile device caching, and it has become an active area of research. In this paper, we first introduce the structure and key concepts of mobile social device caching (MSDC), elaborating the connections and differences of it to traditional caching. Furthermore, we advocate several major areas of research challenges for MSDC, such as content placement, radio resource management and routing. Moreover, a demonstration is present to illustrate the modeling of the MSDC, and performance analysis is carried out to manifest its great potentials.

INDEX TERMS Device caching, mobile social network, physical network, social aware content placement, social aware radio resource management, social aware routing.

I. INTRODUCTION

Mobile data traffic is dramatically increasing due to the explosive growth of mobile devices and services [1]–[3]. Building ultra dense network with the concept of small cell is widely accepted as a key to address the explosive data volume [4], [5]. However, the backhaul from small cell base station to the core network becomes a potential bottleneck for its high cost and limited capacity [6], [7]. With the development of hard disks and individual devices (such as smartphones), [8], large cache space can be provided with comparatively low cost, more and more researchers' attention has been attracted by caching at network edge [1], [9], i.e storing contents in individual devices' caches or in the caches of small base stations (BS). Large amounts of requests can be satisfied locally instead of having to be accessed from the Internet through the backhaul. Especially, caching at individual devices is a promising option with the advent of Device-to-Device (D2D) technology [10] which facilitates cooperation among users.

Traditional caching technologies have been widely used in wired networks, where contents in the main servers are replicated in caching servers near the end users which can greatly reduce service delays, and the burden of the main servers and the overall networks. Currently, most content providers (CPs, such as YouTube) in the Internet adopt the so-called Content Delivery Network (CDN) [6], which is an umbrella term covering complicated enabling technologies that fulfill the aforementioned principle. Despite the popularity and effectiveness, traditional CDN cannot meet the challenges of explosive mobile data volume since it cannot solve the backhaul bottleneck issue mentioned above due to the position of the servers in the network architecture (see Fig.1). Guided with the same principle, researchers have been investigating ways of incorporating caching into the wireless mobile networks architecture, and pushing the caching mechanism to the RAN (Radio Access Network) side. Caching at the base stations or user devices has been an actively researched area.

The concept of Femto-Caching is proposed in [8], where base stations (of femto-cells) are equipment with large volume hard disks serving as content caching servers. Requests can be served directly from base station caches rather than from Internet through backhaul. Not only the access delay

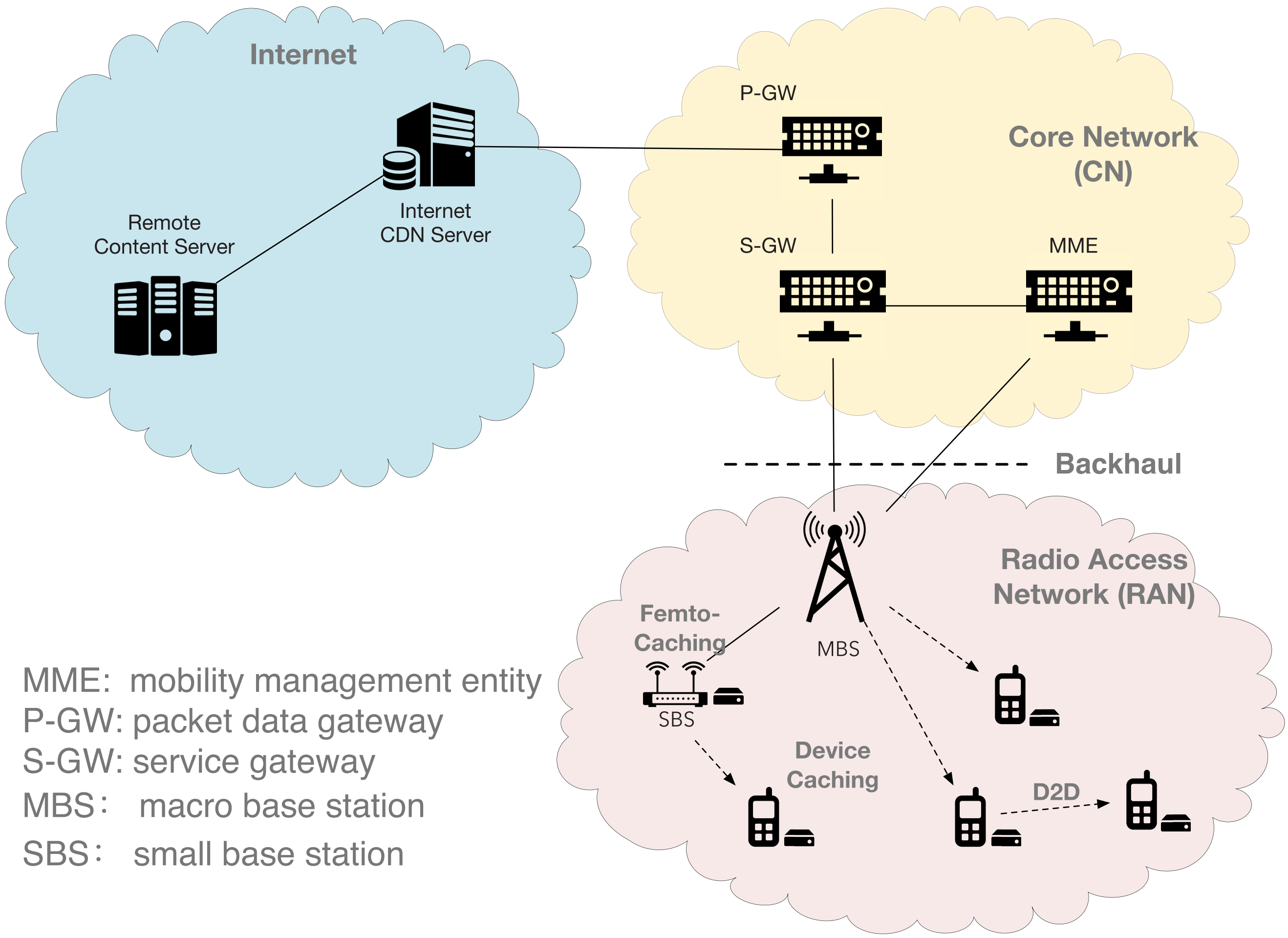


FIGURE 1. An illustration of future wireless network architecture where three types of caching schemes, namely, CDN, Femto-Caching and Device Caching can co-exist.

is further decreased, the burden of Core Network (CN) and backhaul are also significantly alleviated. However, Femto-Caching requires additional infrastructure, i.e the base station of femtocell which is equipped with hard disks. In addition, due to the limited flexibility of BS of femtocell, Femto-Caching can not effectively cope with users distribution especially those areas user gathered.

Besides Femto-Caching, caching at the user devices directly is also an attractive option. With the rapid development of mobile devices in recent years, devices with huge processing power and storage are becoming more common [11]. Traditionally, the performance of device caching is limited by the size of the onboard cache of the device. This issue can be partly alleviated by the fast development of storage capacity governed by Moore's law. Moreover, with advanced strategies such as coded caching [12] or cooperative caching with D2D [1], it can be shown that theoretically the system performance is no longer limited by individual cache size and can scale with the aggregated cache size of all users. Essentially, all user devices create a virtual cache consisting of the sum of the individual caches. Besides, compared to Femto-Caching, device caching is more flexible in structure and can be easily costumed to adapt to the mobility of users. Thus, mobile device caching possesses great potential for solving the problem of massive mobile content distribution.

There is a unique feature that separates caching at the user devices from the other two methods. That is, there are inherent social characteristics that play important roles in the interaction among mobile users. A mobile device is closely connected to its user, and mobile users exhibit strong social characteristics. For example, "close friends" will likely have similar interests and thus their content request patterns might be closely related. Also, "close friends"[1] might meet each other more frequently and have better chance of exchanging via D2D links. Because of the such intimate connection between social characteristics and the interaction of the

[1] Note that here the measure of closeness can be different. The closeness in this example reflects the physical closeness of contact while the closeness in the previous example is more of a measure of the psychological level of closeness

network elements, mobile device caching is social in nature, and we coin the term *mobile social device caching* to better emphasize this feature. Social awareness is essential for the design of the mobile device caching. Specifically, social characteristics exhibit major influences in the following key areas:

1) *Content placement* is always a key issue in caching network which have great impact on system performance [13]. Compared with aforementioned Internet CDN and Femto-Caching, each individual device cache space is limited and accessed opportunistically so that caching performance improvement can be enhanced only through elaborate cooperation among users, e.g., through coded caching or D2D cooperation. User cooperation is inevitably related to the individual user's interest, mobility pattern and willingness to cooperate (friends or not), which are all closely related to the social characteristics of the users.
2) Caching directly in the Radio Access Network (RAN) side is inherently wireless in nature. That means, any practical caching scheme will involve *radio resource management* (RRM), the efficiency of which plays a vital part in system performance. How to better leverage the social characteristics of the network in designing efficient RRM algorithms is a important issue to investigate.
3) For cooperative caching utilizing D2D communication, due to the mobility of individual devices [26] and limited effective communication range [14], multi-hop communication from device to device is a facilitating method for cooperative caching, where *wireless routing algorithms* to construct an end-to-end path between source and destination with multiple intermediate nodes is essential. Considering social characteristic provides a new angel of view in the design of routing protocols which can make better path selection with the help of mobility and willingness pattern of the underlying social network.

In this paper, we will focus on elaborating the research challenges in content placement, radio resource management and wireless routing of mobile social device caching, leveraging the inherent social characteristics of the system. The remained of this paper is organized as follows: We first provide the overall network architecture of mobile social device caching and some key concepts in part II. Part III presents the major research challenges on the design of mobile social device caching. An illustration will be shown to prove the efficiency and potential of mobile social device caching in Part IV. Part V concludes the paper.

II. KEY CONCEPTS OF MOBILE SOCIAL DEVICE CACHING

In this section we introduce a general modeling framework for MSDC and some key concepts within. Naturally we can divide the overall network into two abstract layers, namely, the *social network layer* and the *physical network layer* (see Fig. 2). The social network layer models the social interaction between users of the devices and the physical network layer models the communication mechanism between devices and the supporting infrastructure. Typically, social characteristics of human beings involve mainly (logical) intimacy relationship and (physical) contact relationship among users. Thus the social layer can be further divided into two parts: the online social layer and the offline social layer, with the former reflects the virtual relationships among users and the latter reflects the realistic contact patterns. The two social layers are closely related and interact with each others. In the following, we provide the detailed description of the each layer and introduce some key concepts.

A. SOCIAL NETWORK LAYER

The social network layer provides modeling of the social interaction between users of devices. It can be further divided into two kinds, namely, the online and offline social networks. The online social network mainly focuses on the virtual relationship among users while offline social network pays more attention to users' realistic mobility. There are also close connections between these two networks. First, we introduce some common concepts of social networks which is applicable to both layers.

- *Social Tie and Tie Strength:* Social tie is the most basic concept of a social network, the strength of which characterizes the extend by which two individuals are related to each other. In a social network, user A is related to user B if there is a social tie connecting each other. The strength of a social tie measures quantitively the quality of the corresponding social relationship, and can be used to also qualitatively define common relationships such as friends and strangers. In online social networks, social tie reflects the intimacy relationship of the users, which can be used to model the interest similarity, e.g., friends will have more similar interests than strangers. In offline social networks, social tie can be used to model the contact frequency of users' mobility, and friends will make contact more frequently than strangers. In MSDC, social tie and tie strength of online social networks can be used for modeling request pattern relationship of users and those of offline social networks can be used for modeling the mobility pattern.
- *Community:* Individuals of a social network can be clustered based on social tie strength and form communities. A community is a structure sub-unit (which can be represented as a set of individuals) of a social network with high density of internal links. Individuals have more social connections with other individuals inside their own community than with individuals outside (see an illustration in Fig. 2). As mentioned, usually, users within the same community of the online social network may have similar interests and have large probability of requesting same contents. Likewise, the users from

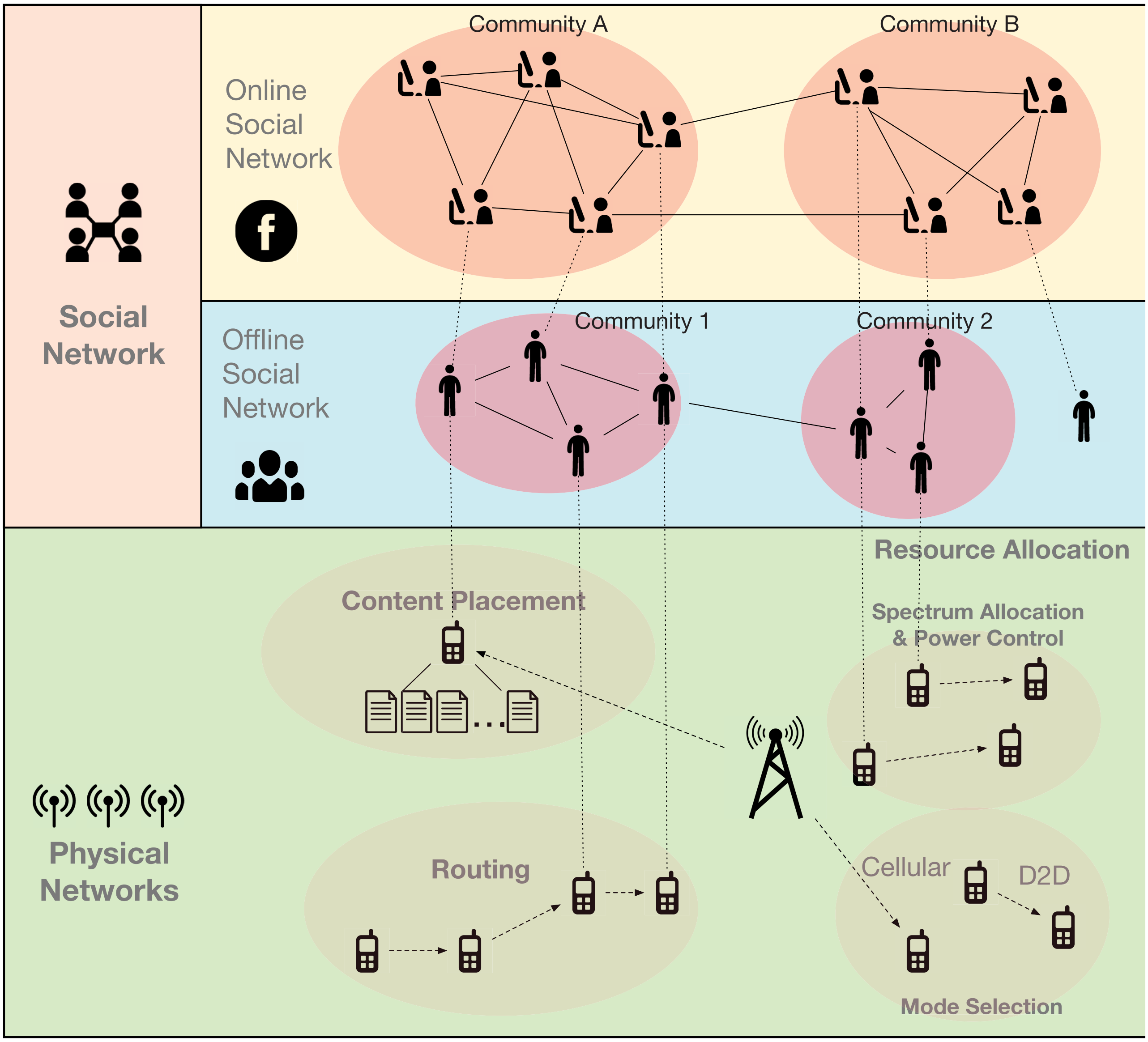


FIGURE 2. The layered structure of mobile social device caching.

the same community of offline social network will meet each other more often.

- *Centrality:* In human society people possess heterogeneous popularity [16]. Centrality is used to describe such heterogeneity in social networks, the score of which can denote a user's importance. High centrality individuals have more social ties (degree centrality) or larger capacity of connecting others (between centrality and closeness centrality) [18]. In online social networks, a user of high centrality has high influence on others' interests. For example, the contents requested by such a user are more likely to be requested by other users in the community. In offline social networks, a user of high centrality will meet more people and thus can be a good media of content sharing.

Furthermore, we introduce some key concepts that are unique for the respective online and offline layers of the social networks.

- *Selfishness:* Users with selfishness are most likely not willing to help others while sacrificing their own limited resources (such as battery life, cache space and wireless spectrum) and make decision aiming at maximizing their own utility. Selfishness is a salient characteristic in online social network and presents users' willingness for serving others which has great impact on system design. There are two types of selfishness, individual and social selfishness [16]. Users with individual selfishness treat others equally while users with social selfishness have different levels of selfishness to others based on tie strength or community.

- *Inter Contact Time:* Inter contact time reflects the time between two contacts of a offline social network. It is commonly assumed to follow an exponential distribution and contacts between two users can be formulated as a Poisson process with parameter λ. Generally, individuals equipped with strong social ties or from same community may have large λ.
- *Contact Duration:* Contact duration reflects the duration of one contact between two nodes of a offline social network and will determine the communication time of a D2D link. In [12], it shows that the contact duration typically follows a Pareto distribution. The length of contact duration will have great impact on the amount of data successfully transmitted during one contact so that it will effect content placement and relay node selection in routing.

B. PHYSICAL NETWORK

The physical network represents the user devices and the related supporting infrastructure such as base stations. In the physical network we considered, there are multiple devices in BS covered area where each device also corresponds to one user in online social network and one user in offline social network. Each device is equipped with certain onboard storage as cache which is small in size compared to those in CDN or Femto-Caching, such that only limited amount of contents can be cached. The devices support two communication modes, namely, the cellular mode linking to the BS and the D2D mode linking with users nearby. When accessing a content file, a device first search the file in its local cache. If it is not present, it search over the others devices' cache nearby. Finally, it will request the file from the BS if the request is not met by caching. Due to the limit of onboard caching, user cooperation is vital in MSDC and an well-designed content placement scheme is needed to increase the efficiency of cache sharing. Further, interference shall be managed in case of D2D communications in which radio resource management schemes should be implemented (mainly involves spectrum allocation, power control, mode selection). Moreover, due to the limited range of D2D communications, to further increase the sharing of caching, multi-hop D2D links can be established where (ad-hoc) routing algorithms are needed.

III. CHALLENGES IN MOBILE SOCIAL DEVICE CACHING

In this section, we address the major research challenges in mobile social device caching in the areas of 1) *content placement*: where, which and when contents should be cached in multiple individual device; 2) *radio resource allocation*: how to construct D2D or cellular links to communicate and share which subject to physical resource constraint; and 3) *routing*: how to effectively construct multiple D2D links from source cache to destination to satisfy users' requests locally. We will argue how social network and its characteristics can play a vital role in improving the efficiency of MSDC in these areas.

A. SOCIAL-AWARE CONTENT PLACEMENT

Content placement is always a key issue in any caching-related schemes. The crucial question is succinctly summarized as *which,where and when content should be cached*. In mobile social device caching, because of the primary use of D2D short-range communication and the opportunistic contact among mobile users, content access in device cache is always opportunistic in nature. In addition, due to the limited contact duration and spectrum, only limited amount of contents can be transmitted during one contact. Furthermore, stringent battery budget and limited individual device cache space makes the message flooding impractical and unacceptable in MSDC [13]. Therefore, effective content placement becomes an even more crucial issue to the effectiveness of the system in MSDC.

1) REACTIVE v.s. PROACTIVE

Reactive content placement means that the requested contents are cached immediately at the time of the arrival of the request, which is typically when the network is at peak hour traffic. Therefore reactive mechanism can't effectively cope with peak traffic overload situation. To smooth out the network traffic, predictable peak hour contents can be transmitted and cached *proactively* during off-peak hour which both reduce the peak hour traffic and eliminate download waiting time [1]. Obviously, the precision of future request prediction determines the performance of proactive content placement. Social characteristic are proved to have strong correlation to request patterns and can contribute greatly to request prediction. For example, users have large probability of requesting those contents which have been requested by other users in the same community or by other users sharing strong social ties (i.e., friends). Therefore, social characteristics can help predicting future request in MSDC. In reality, however, many contents are released before peak hour when the network load is comparatively heavy and those contents may be requested frequently in the near future. How to rapidly and precisely make proactive decision under the consideration of current tasks and future requests can be a challenge.

2) COOPERATIVE v.s. SELFISH

Most of the users in MSDC are selfish in nature since the usage of the devices are limited by battery capacity and storage size. Selfish users aim at maximizing their own utility while avoiding contributing the limited resource to others. Such selfish behaviors may deviate from network-wide optimal policy and significantly degrade the system performance since for MSDC to work well, user cooperation is the key. Therefore incentive mechanisms need to be designed to motivate users to contribute their own resource. Currently there are four basic incentive mechanisms as follows: a) *reputation-based*: users gain reputation if they share their cached contents with others. For rewarding, high reputation users' interested contents gain high weight and are more likely to be cached; b) *credit-based*: users can earn credit to share contents with others and can spend credit to access contents from other users; c) *Tit-for-TAT*: this idea is based on providing and caching equal amount of contents for each other; d) *social-trust based*: users place and share contents among friends while not for others [16], [18].

However, level of users' selfishness is not invariant, which is usually influenced by current mode and the status of its resource. For example, those user, carrying devices with fully charged battery in idle mode, are more willing to contribute their resource to help others. Therefore, the detection and utilization of variable selfishness is an opening challenges for content placement.

3) CENTRALIZED v.s. DISTRIBUTED

It is obvious that if a central controller determines all users' content placement according to the global information, such as movement pattern, request pattern, and the full channel state information (CSI), it might be possible to make the best placement decision. However, full CSI is scarce in practice [21]. Moreover, centralized content placement also causes overhead and imposes a heavy computational burden on central controller, which can become impractical in large networks, especially for MSDC with increasing number of users. Therefore, decentralized content placement schemes should be designed and the concept community from social networks plays a great role. Leveraging the community structure can simplify the CSI detection [15] and therefore help make decentralized content placement decision more quickly and accurately. In addition, due to the nature of community where dense relationships exist inside a community and sparse relationships between different communities, users can be divided into multiple independent groups. Significant performance can be achieved under the consideration of information within community like in [13]. In reality, some users may belong to multiple communities and have probability of making different decisions in different communities. On this occasion, there exits a game among different communities which can lead to many interesting research topics.

4) UNCODED v.s. CODED

Traditional CDN and Femto-Caching have comparatively large cache space that contents can be cached in full which is usually called as *uncoded* caching. However in MSDC, each individual device cache has limited space and thus can only store few number of integral contents. In addition, due to the opportunistic contact between users, the integral content may not be transmitted successfully during one contact. Therefore, contents can be divided into multiple fragments through effective *coding* schemes and distributed to several users [27]. Usually, coded content placement is designed carefully such that coded multicasting opportunities are available which can effectively reduce peak network load and outperform uncoded content placement [12]. Social characteristics can also make coded content placement more effective. For example, those fragments can be placed in devices with high centrality for the convince of sharing. Although coded caching outperforms uncoded caching in many aspects, caching schemes usually generate pressures on computing capacity, especially for mobile devices. Effective coding schemes with low complexity should also be considered. Therefore, how to find the right balance between performance and complexity should be a great challenge for MSDC.

B. SOCIAL-AWARE RADIO RESOURCE MANAGEMENT

In MSDC, D2D links are established based on content demand and potentially shares the same spectrum with cellular links which inevitably causes interference issue. Radio resource management strategies are thus important to mitigate interference and to ensure both D2D and cellular link quality. There are two main areas of radio resource management, namely, resource allocation and mode selection, that we think are to be well adjusted to accommodate limited resource and users' mobility in MSDC.

1) RESOURCE ALLOCATION

In MSDC, resource allocation mainly involves spectrum allocation and power control which allocate spectrum and transmit power for those links to mitigate interference and ensure communication quality [15]. Due to the limited duration of each D2D link and stringent battery budget, throughput for D2D links and energy efficiency for users should be considered to ensure successful content transmission and power saving [19].

Guided with related social characteristics, for high centrality users, they have more chance to share cached contents with others through D2D links which however can easily leads to traffic congestion (for its own need) and large amount of power consumption. In this case, careful spectrum allocation scheme should be designed to avoid such congestion and effective resource allocation and power control mechanism should also be considered to save power so as to prolong the lifetime of the devices.

For typical cellular users, due to the nature of selfishness, they are more willing to share their cellular spectrums with others in the same community [17]. In addition, users in the same community will act as a whole and try to maximize the community utility. However, users in the same community are usually geographically close to each other, whereby serve interference becomes an issue if they share the same spectrum. More intelligent resource allocation should be designed to take community information into account [17]. For example, for D2D links, users in one community can use the cellular spectrum used by a community that is located far way. Thus spatial frequency reuse is applied to a broader sense of communities.

2) MODE SELECTION

In MSDC, each user can access contents from BS with cellular link or from other users with D2D link. Due to the same interface shared by the two communication links, usually only one communication mode is supported at the same time. Mode selection decides which communication mode to choose and which user to select as content provider if D2D communication mode is selected.

Generally, for high centrality users, they should have high priority to link to BS to access more contents since they have large chance to contact others and share contents with others. In addition, when multiple potential content providers exists, those high centrality users should have comparatively lower priority to serve than common users so as to relieve their work load and save their power. Meanwhile, those users with strong ties are more likely to be selected as content provider due to the reason that users with strong ties may have a longer contact duration which can ensure successful content transmission [12].

C. SOCIAL-AWARE ROUTING

Due to the users' mobility and limited D2D communication range in MSDC, D2D links are opportunistic and contents are not always accessed within one hop scope. We can therefore construct an end-to-end path with multiple intermediate nodes to render those contents accessible. The mainstream approach is the so-called "store-carry-forward" approach, which means that nodes store and carry the contents until they encounter other suitable nodes [18]. Relay node selection is major issue in routing design for MSDC and mobility prediction is a key challenge. In MSDC, devices are carried by human with social characteristics, with the help of which opportunistic contact prediction can be made more precisely. For example, with the help of social properties such as community and centrality, the sender can first make effort to send content to any users belonging to the destination community. Then, content is sent to user with high centrality in this community who have more probability of contacting others within the same community. And eventually the high centrality user can transmit the content to the destination, with higher probability than others [22].

However, community information is not always accessible due to difficulty of community detection. Besides community information, other social characteristics can also help the design of routing mechanisms. For example, with the help of tie strength, centrality, and similarity which are introduced in [16], effective routing algorithms can also be proposed. Such algorithms avoid difficult community detection problems and can be easily implemented.

All routing mechanisms mentioned above are based on the assumption that all users are willing and altruistic to act as relay nodes. In addition, there are also many incentive approaches to overcome the potential selfishness issue (individual selfishness and social selfishness) and stimulate cooperation among users while basic ideas have been proposed above.

Meanwhile, social aware routing also puts users and network at risk, especially for more open environment, such as Internet of things [23]. This would cause safety problems and generate great pressure mainly on three aspects: a) *Trust*: Trust in MSDC is defined as the willingness of a particular user which is acquired slowly and destroyed quickly [24]. And trust establishment and maintenance in MSDC is always an open issue. b) *Security*: The purpose of security is to protect the information and the resources against attacks. Recognized protocols are urgent needed to be designed to defend multiple attacks [25]. c) *Privacy*: Usually, more information sharing (such as social ties, cached contents and mobility information) can improve routing efficiency while also leads to private information leakage risk [18]. Therefore, finding the most optimal tradeoff between privacy and efficiency in MSDC will be the challenge in the future.

The performance of most existing social aware routing protocols depends on stable and accurate social characteristics [16]. However, facing time-varying topography, only dynamic and less accurate social characteristics are accessible. Therefore, utilizing those social characteristics to design routing protocols becomes an open challenge and needs to be further investigated.

IV. SOCIAL-AWARE CACHING VIA D2D: AN ILLUSTRATION

In this section, we present a concrete example to demonstrate the potential of utilizing social characteristics to improve the performance of content placement in MSDC. We consider a wireless system consisting of a single base station and M mobile users. The base station is connected to a remote file server via wired/wireless backhaul. The users are interested in accessing the files stored in the remote file server through the base station. Each user device has a local cache of fixed size and can cache part of the content library. Meanwhile, if two users are in proximity, D2D communication can be established for file sharing. We assume that each pair of user from same community have same request probability for the contents while users from different community have different request probability. Additionally, each pair of users from same community have same contact probability which is larger than those from different communities (friends contact more often than strangers).

A caching scheme in discussion can be divided into two phases, namely, the content placement phase and the content delivery phase. For content delivery, at any moment, user requests a file according to its popularity distribution. The file is accessed from the local cache of the device if it is present. Otherwise, it is accessed from the users in its proximity if it is present in any of their local cache via D2D transmission. If the file is not present in the local cache of the user or its neighbors, it is transmitted by the remote file server via the base station. Our goal here is to find smart content placement schemes that minimize the chance of requesting files from the remote server. We define the probability that user m can access file f from its local cache or that of its encountered users as the *average hit ratio*, which can be written as:

$$\frac{1}{M}\sum_{m\in\mathcal{M}}\sum_{f\in\mathcal{F}}\left(1-\prod_{m'\in\mathcal{M}}\left(1-c_{m',f}\,q_{m,m'}\right)\right)p_{m,f}, \quad (1)$$

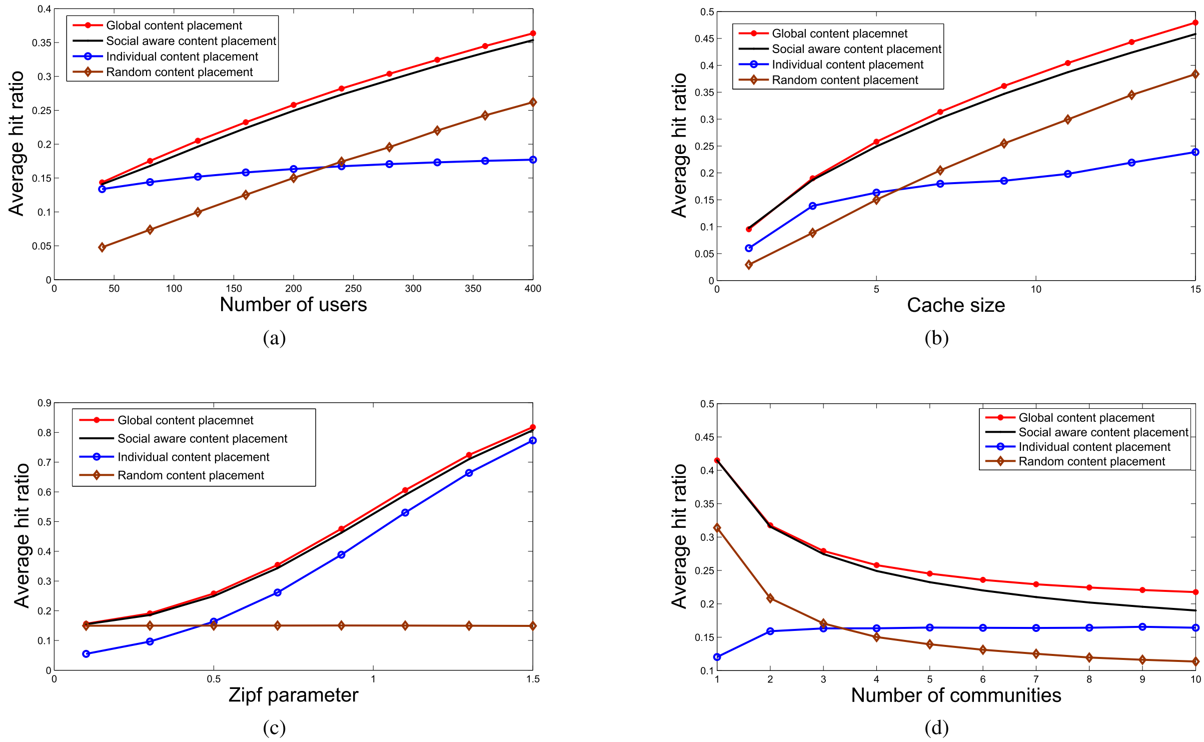


FIGURE 3. Simulation results for the four proposed content placement mechanisms. a Average hit ratio as number of users changes. b Average hit ratio as cache size changes. c Average hit ratio as Zipf parameter α changes. d Average hit ratio as number of communities changes.

where $\mathcal{M}$ represents the user set, $\mathcal{F}$ represents the content library, $c_{m,f}$ shows whether the user m stores content f, $p_{m,f}$ is the request probability of user m for content f, and $q_{m',m}$ indicates the contact probability between user m' and user m where two users are in proximity to establish different D2D links for file sharing. It can be proved that (1) is a monotone *submodular* function and we can apply *semigradient-based maximization* algorithm which has lower complexity and converges much faster than the classic algorithm introduced in [8]. Please refer to [20] for the detailed caching algorithm and simulation settings.

Here, we propose four caching mechanisms based on the information utilized when placing contents. *Global content placement* makes content placement decision based on the all users' request and contact probability information. *Social aware content placement* places content based on the users' information within their respective community and there is no information exchange among different communities. *Individual content placement* indicates that user places contents based on their own interest only. Contents are randomly placed in *random content placement*.

Fig. 3 (a) and (b) depict the simulation results of average hit ratio as the user number and cache size increase. As user number or cache size increases, more content copies can be cached and shared among users locally so that hit ratio grows. In Fig.3 (c), hit ratio grows with larger Zipf parameter α. Larger Zipf parameter α means steeper request probability distribution where majority of requests probably concentrate on the limited number of high rank popular contents. Fig.3 (d) shows the evaluation of hit ratio as the community number changes. As community number grows, users' interest becomes different and there is smaller probability of content reusing. However, for individual content placement, there are more different contents stored and larger probability of sharing as community number increases.

Additionally, though the comparison of different caching schemes in Fig. 3, we can observe that as the average hit ratio scales nicely with global and social-aware caching. Random caching is proved to scale optimally in some asymptotically case but here we observe that in practical situation it has a almost contact gap with the global and social-aware caching. Individual caching without utilizing social information performs worst and does not scale nicely as the number of users increases. Thus, we can see that social information can greatly benefit the system performance of MSDC. Meanwhile, we observe that social aware caching achieves almost identical performance to the global caching scheme, which indicate that local information within a community might suffice to guarantee such performance gain.

V. CONCLUSION

In this paper, we have introduced the concept of mobile social device caching and addressed several major research challenges in the areas of content placement, radio resource management and routing. The key is how to effective incorporate the social characteristics that is naturally inherent in MSDC into the design of such networks and this is reflected in the social-awareness design guidelines proposed. Through an example, we have demonstrated the great potential in MSDC and shows that MSDC can be a promising way to meet the stringent demand of future mobile data era.

REFERENCES

[1] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.
[2] C.-X. Wang *et al.*, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.
[3] R. Zi, X. Ge, J. Thompson, C.-X. Wang, H. Wang, and T. Han, "Energy efficiency optimization of 5G radio frequency chain systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758–771, Apr. 2016.
[4] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sep. 2008.
[5] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 36–43, May 2014.
[6] H. Ahlehagh and S. Dey, "Video-aware scheduling and caching in the radio access network," *IEEE/ACM Trans. Netw.*, vol. 22, no. 5, pp. 1444–1462, Oct. 2014.
[7] 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.
[8] N. Golrezaei, A. F. Molisch, A. G. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 142–149, Apr. 2013.
[9] 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.
[10] C. Xu, C. Gao, Z. Zhou, Z. Chang, and Y. Jia, "Social network-based content delivery in device-to-device underlay cellular networks using matching theory," *IEEE Access*, pp. 1–12, Nov. 2016.
[11] Z. Chang, Y. Gu, Z. Han, X. Chen, and T. Ristaniemi, "Context-aware data caching for 5G heterogeneous small cells networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
[12] X. Zhuo, Q. Li, G. Cao, Y. Dai, B. Szymanski, and T. La Porta, "Social-based cooperative caching in DTNs: A contact duration aware approach," in *Proc. IEEE 8th Int. Conf. Mobile Ad-Hoc Sensor Syst.*, Oct. 2011, pp. 92–101.
[13] Y. Wang, J. Wu, and M. Xiao, "Hierarchical cooperative caching in mobile opportunistic social networks," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 411–416.
[14] Y. Yang, C. Zhao, S. Yao, W. Zhang, X. Ge, and G. Mao, "Delay performance of network-coding-based epidemic routing," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3676–3684, May 2016.
[15] Y. Li, T. Wu, P. Hui, D. Jin, and S. Chen, "Social-aware D2D communications: Qualitative insights and quantitative analysis," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 150–158, Jun. 2014.
[16] F. Xia, L. Liu, J. Li, J. Ma, and A. V. Vasilakos, "Socially aware networking: A survey," *IEEE Syst. J.*, vol. 9, no. 3, pp. 904–921, Sep. 2015.
[17] F. Wang, Y. Li, Z. Wang, and Z. Yang, "Social-community-aware resource allocation for D2D communications underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3628–3640, May 2016.
[18] Y. Zhu, B. Xu, X. Shi, and Y. Wang, "A survey of social-based routing in delay tolerant networks: Positive and negative social effects," *IEEE Commun. Surveys Tut.*, vol. 15, no. 1, pp. 387–401, 1st Quart., 2013.
[19] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tut.*, vol. 16, no. 4, pp. 1801–1819, 4th Quart., 2014.
[20] Y. Wu, S. Yao, Y. Yang, Z. Hu, and C. Wang, "Semigradient-based cooperative caching algorithm for mobile social networks," in *Proc. IEEE Global Commun. Conf.*, pp. 1–6, 2016.
[21] L. Marini, J. Li, and Y. Li, "Distributed caching based on decentralized learning automata," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 3807–3812.
[22] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE rap: Social-based forwarding in delay-tolerant networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 11, pp. 1576–1589, Nov. 2011.
[23] Q. Xu, P. Ren, H. Song, and Q. Du, "Security enhancement for IoT communications exposed to eavesdroppers with uncertain locations," *IEEE Access*, vol. 4, pp. 2840–2853, 2016.
[24] Y. Najaflou, B. Jedari, F. Xia, L. T. Yang, and M. S. Obaidat, "Safety challenges and solutions in mobile social networks," *IEEE Syst. J.*, vol. 9, no. 3, pp. 834–854, Sep. 2013.
[25] Y. Sun, L. Yin, and W. Liu, "Defending sybil attacks in mobile social networks," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr./May 2014, pp. 163–164.
[26] X. Ge, J. Ye, Y. Yang, and Q. Li, "User mobility evaluation for 5G small cell networks based on individual mobility model," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 528–541, Mar. 2016.
[27] 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.

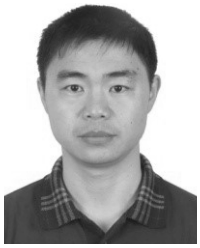
YECHENG WU received the B.S. degree from the Department of Optical Information Science and Technology, Nanjing University of Post and Telecommunications, Nanjing, China, in 2013. He is currently pursuing the Ph.D. degree with the Shanghai Institute of Microsystem and Information technology, Chinese Academy of Sciences, Shanghai, China. His research interests include wireless caching, social network analysis, mobile edge computing, and resource allocation in fifth-generation networks.

SHA YAO received the B.Sc. degree from Zhejiang University, Hangzhou, China, in 2004, and the M.Sc. and Ph.D. degrees from the Royal Institute of Technology, Stockholm, Sweden, in 2006 and 2011, respectively, all in electrical engineering. His research interests include information-theoretic perspectives on cooperative transmission and relay networks, resource allocation in fifth-generation heterogeneous networks, and task scheduling in crowdsensing.

YANG YANG received the B.Eng. and M.Eng. degrees in radio engineering from Southeast University, Nanjing, China, in 1996 and 1999, respectively, and the Ph.D. degree in information engineering from The Chinese University of Hong Kong in 2002. He has served with the Department of Electronic and Electrical Engineering, University College London, U.K., as a Senior Lecturer; the Department of Electronic and Computer Engineering, Brunel University, U.K., as a Lecturer; and the Department of Information Engineering at The Chinese University of Hong Kong as an Assistant Professor. He is currently a Professor with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, where he is also serving as the Director of the CAS Key Laboratory of Wireless Sensor Network and Communication, and the Director of the Shanghai Research Center for Wireless Communications. He is also an Adjunct Professor with the School of Information Science and Technology, ShanghaiTech University. He has co-edited the book *Heterogeneous Cellular Networks* (Cambridge University Press, 2013) and co-authored over 100 technical papers. His research interests include wireless ad hoc and sensor networks, software defined wireless networks, 5G mobile systems, intelligent transport systems, wireless testbed development, and practical experiments.

TING ZHOU was born in Guangdong in 1982. She received the B.S. and M.S. degrees from the Department of electronic engineering, Tsinghua University, in 2004 and 2006, respectively, and the Ph.D. degree from the Shanghai Institute of Microsystem and Information Technology (SIMIT), Chinese Academy of Sciences, in 2011. From 2011 to 2013, she was with the Shanghai Research Center for Wireless Communication, SIMIT, as a Research Assistant. Since 2014, she has been an Associate Professor with the Shanghai Research Center for Wireless Communication, SIMIT. She has authored or co-authored over 20 papers published in journals and conferences, and 44 granted and pending patents. Her current research interests include resource management and intelligent networking of heterogeneous wireless networks. She received the 2015 First Prize of China Institute of Communication Technical Innovation Award and the 2015 Second Prize of Shanghai Science and Technology Progress Award.

HUA QIAN received the B.S. and M.S. degrees from the Department of Electrical Engineering, Tsinghua University, in 1998 and 2000, respectively, and the Ph.D. degree from the School of Electrical and Computer Engineering, Georgia Institute of Technology, in 2005. He is currently with the Shanghai Advanced Research Institute, CAS, as a Full Professor. He has co-authored two book chapters, published over 60 peer-reviewed papers, and applied for over 50 patents with four granted U.S. patents. His current research interests include nonlinear signal processing and system design of wireless communications.



HONGLIN HU (SM'13) received the Ph.D. degree in communications and information system from the University of Science and Technology of China in 2004. Since 2016, he has been with the Shanghai Advanced Research Institute, Chinese Academy of Sciences. He received the 2016 IEEE Jack Neubauer Memorial Award. He was the Vice Chair of the IEEE Shanghai Section. He was the leading Guest Editor of the IEEE Wireless Communications Special Issue on Mobile Converged Networks, and the *IEEE Communications Magazine* Special Issue on Software Defined Wireless Networks (Part I and Part II). He is a Finland Distinguished Professor. He serves as an Associate Editor for four international journals.

MATTI HAMALAINEN was a Senior Researcher with the Helsinki University of Technology from 2001 to 2006 and a Professor with Aalto University from 2006 to 2011. He joined DIGILE in 2011 as a Coordinator of the China-Finland Strategic ICT Alliance and continued at DIMEEC when DIGILE and FIMECC merged in 2016. He also serves as a National High-end Foreign Expert with Tongji University, Shanghai, and the Director of the ISeeLife! Lab-International Open Lab on Intelligent Services Embedded in Everyday Life, Sino-Finnish Center, Tongji University.