Energy Efficiency Analysis of Heterogeneous Cache-enabled 5G Hyper Cellular Networks

Jiaxin Zhang*†, Xing Zhang*, Muhammad Ali Imran†‡, Barry Evans† and Wenbo Wang*
*Wireless Signal Processing & Networks Lab (WSPN)
Beijing University of Posts and Telecommunications (BUPT), Beijing, China
Corresponding Author: Jiaxin Zhang, email: zhangjxbupt@gmail.com
† Institute for Communication Systems (ICS)
University of Surrey, Guildford, UK
‡ University of Glasgow, Scotland, UK

***Abstract*—The emerging 5G wireless networks will pose extreme requirements such as high throughput and low latency. Caching as a promising technology can effectively decrease latency and provide customized services based on group users behaviour (GUB). In this paper, we carry out the energy efficiency analysis in the cache-enabled hyper cellular networks (HCNs), where the macro cells and small cells (SCs) are deployed heterogeneously with the control and user plane (C/U) split. Benefiting from the assistance of macro cells, a novel access scheme is proposed according to both user interest and fairness of service, where the SCs can turn into semi-sleep mode. Expressions of coverage probability, throughput and energy efficiency (EE) are derived analytically as the functions of key parameters, including the cache ability, search radius and backhaul limitation. Numerical results show that the proposed scheme in HCNs can increase the network coverage probability by more than 200% compared with the single-tier networks. The network EE can be improved by 54% than the nearest access scheme, with larger research radius and higher SC cache capacity under lower traffic load. Our performance study provides insights into the efficient use of cache in the 5G software defined networking (SDN).**

***Index Terms*—Cache-enabled Networks, HCNs, SDN, C/U Split, Energy Efficiency**

I. Introduction

The next generation networks (5G) are expected to support an increasing number of connected devices and diversity of applications, which requires wireless communication systems to move towards a real information-and-user based network. Thanks to the prevalence of the on-line social websites and instant messaging systems, the study of social network research in temporal, spatial and content domains have been utilized to characterize the group user behaviors [1][2][3]. Deploying small cells can help to satisfy high-speed requirements and overcome the bottlenecks of limited backhaul. In work [4][5][6], the request probability of a particular content is modeled as Zipf distribution, and the most popular contents are cached in the network. In [5], the single tier network is studied to derive the impact of backhaul capacity and size of cache on the energy efficiency. In [6], the performance of the wireless heterogeneous network where the radio access network caching and D2D caching coexist is studied and comparisons are made between cache-enabled network and the system without caching ability.

However, all of these strategies are based on a snap-shot of the network, the traffic fluctuation as well as the group user behaviors in temporal and spatial domain are neglected. With more and more SCs deployed in future, the traffic fluctuation will be more obvious so that the traffic requirements are non-uniformly distributed. The always-on cell configuration brings huge power waste in the network. In addition, the fairness of users is neglected, because only the most several popular files are cached in the nearby SCs in these works. In [7] a caching scheme to determine which replicas should be stored in content store is proposed. In work [8] the files in small cells are randomly selected and cached, assuming that the popularity of several cached contents follows a uniform distribution. These works help to improve the fairness of users and can be adopted in dynamic of traffic requirements.

In order to improve the manageability and adaptability of the network, the control and data plane separation based on the software orchestration mechanism has attracted considerable attention in both industries and academics [9] [10]. In this paper investigate the caching problem in hyper cellular network. We focus on the utilize of the control and user plane separation in the cache-enabled networks. Different from the traditional cache-enabled network, with the software-defined features in this network, the C/U planes are separated to improve the manageability and adaptability of the network. The macro cells, with the master Radio Resource Management (RRM) function, play important roles in both C-plane and U-plane, while the SCs only provide slave-RRM function in the U-plane. Equipped with different protocols, the macro cells are in charge of the mobility connection, home subscriber server and access admission control for the users, and the SCs are designed toward pure data-only base stations with the assistance of macro cells. In this software defined networks, the public overhead cost can be reduced, and the network resources can be scheduled in a more efficient way.

Compared with the traditional Content Delivery Network(CDN), cache strategies in the HCNs with software-defined features are different in the following aspects:

- The user traffic requirements vary in both temporal and spatial domain. So that the resource scheduling method and on-off schemes can be developed according to the

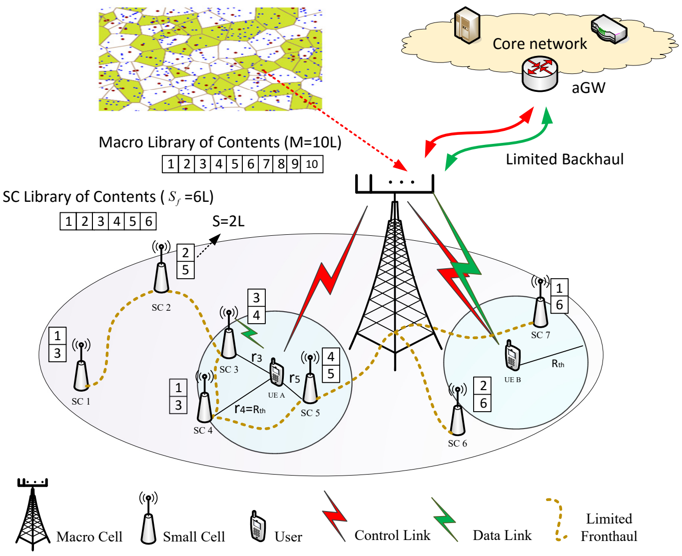

Fig. 1. Illustrate of the heterogeneous cache-enabled 5G software defined networks with the limitation of backhaul to the core network. The cache abilities of macro cells and SCs is M=10L and S=2L respectively. UE A obtains access to SC3 and UE B obtains access to the macro cell.

traffic fluctuation.

- The macro cells and SCs play different roles in the software defined networks, including the cache ability, on-off status and backhaul limitation.
- Not only the content hit probability and throughput of the network, but also the computing and delivery energy consumption, and the access and user association strategy should be considered emphatically.

Given these differences, the aim of this work is to investigate the caching strategy in hyper cellular network. We also propose a novel access scheme under the C/U split architecture based on user interest and fairness of service, where the on-off strategy of SCs can be adopted based on the traffic fluctuation. The coverage probability and EE are derived theoretically and comparisons are made with the single-tier system and nearest access scheme. Numerical results show that the coverage performance can be enhanced by more than 200% and the EE can be improved by 54% with larger research radius and higher SC cache ability under lower traffic condition.

The rest of this paper is organized as follows: Section II describes the proposed system model of the cache-enabled HCNs and a novel access procedure is proposed. The probability of hit, service coverage probability, throughput, energy consumption and EE are derived study the influence of various key parameters in Section III. Simulation results are obtained and analysed in Section IV. Finally, conclusions and future works are drawn in Section V.

II. System model

We consider the downlink transmission in HCNs, where the BSs are distributed on the two-dimensional Euclidean plane $\mathbb{R}^2$, illustrated in Fig. 1. Based on the stochastic geometry theory, macro cells and SCs are modelled as independent two-tier homogeneous Poisson Point Processes (PPPs), denoted as Φ_M and Φ_S, with corresponding densities of λ_M and λ_S. The users are positioned with PPP Φ_U with the density of λ_u.

A. Pathloss and Channel Model

The standard path loss propagation model is used with path loss exponent $\alpha > 2$. The identical Rayleigh fading channel model is used independently on all links as exponential distribution with mean of 1: $h_u^M \sim \exp(1)$, $h_u^S \sim \exp(1)$. We assume the heterogeneous networks is an interference-limited scenario, so that the constant additive noise is neglected. Without loss of generality, a mobile user u is chosen to be located at the origin, and the distances to the serving macro cell and SC are defined as r_{Mu} and r_{Su} respectively. The downlink signal-to-interference ratio (SIR) from the associated macro cell or the SC is modelled as:

$$SIR_M = \frac{P_M^t h_u^M r_{Mu}^{-\alpha}}{\sum\limits_{M' \in \Phi_M} P_M^t h_u^M {r'_{M'u}}^{-\alpha}}, \tag{1}$$

$$SIR_S = \frac{P_S^t h_u^S r_{Su}^{-\alpha}}{\sum\limits_{S' \in \Phi_S} P_S^t h_u^S {r'_{S'u}}^{-\alpha}}, \tag{2}$$

where P_M^t and P_S^t are the transmission power of the macro and small cells receptively.

B. Cache Deployment Model

In this paper, the trunks of contents requested by most of the terminals can be pre-loaded and stored in both macro cells and SCs, which can help to reduce the latency for UE to obtain access and lower down the power consumption. Denote that the total number of files in the content library is N_f and each file of length of L. According to [4][5][6], the file popularity can be modelled as Zipf distribution, and the probability of the $f-th$ ranked content is requested by terminals is

$$P_{Nf}(f) = \frac{f^{-v}}{\sum_{k=1}^{Nf} k^{-v}}, \tag{3}$$

where $v \geq 0$ reflects the skew of the popularity distribution. The sizes of SC and macro cell library are S_f and M respectively. Due to the fact that different group of people have totally different behavior and specific interest for the contents, from the consideration of fairness, the SCs, with the cache size of S are assumed to store several most popular files randomly from the library. In this way, the probability of hit and fairness are considered simultaneously. For instance, shown in Fig. 1, the macro cell caches most popular $M = 10$ files in the library. Each SC randomly selects and caches $S = 2$ files uniformly from the given SC library $S_f = 6$.

C. User Association

In this subsection, we propose a novel random access strategy based on the software-defined features. Different from current LTE networks, a typical user in the HCNs acquires the basic system information by receiving the master system information from the macro cell initially and then obtains access to the nearest macro cell in C-plane. After that, the UE sends content requests within a certain region with a search

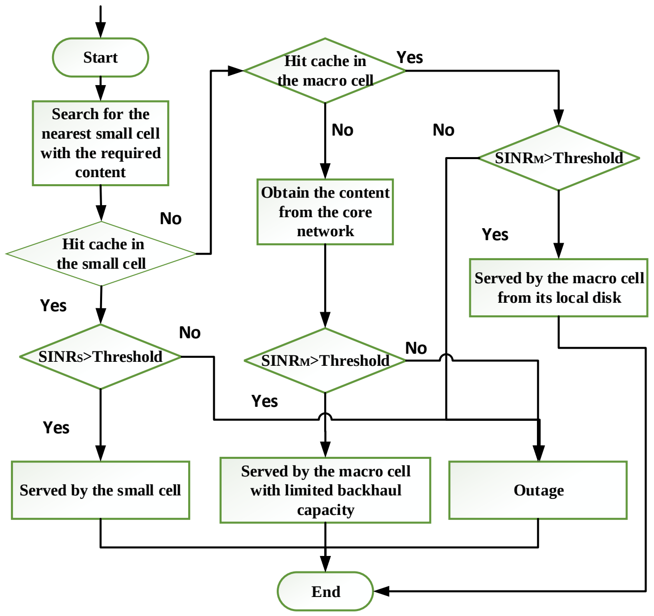

Fig. 2. A novel procedure of random access in the cache-enabled software defined networks.

distance threshold of R_{th}, from the aspects uplink of power control and energy saving.

With the assistance of macro cell, SCs within the region can work in semi-sleep mode to listen to the uplink signals from the user and send the reports back to macro cell via fronthaul connection on Xn interface. Thus the macro cells know the overall information about the SC on-off status, cache library lists and distances to the UE. The macro can deicide which SC or itself provide service to the user. If the small cell is the best serving cell, the macro cell can trigger the access procedure and turn on this SC. After providing the service to the UE, the SC then turns into the sleep mode immediately. As shown in Fig. 2, in the cache-enabled software defined networks, the access decision procedure in U-plane the following procedure:

1) If the cache in one of the SCs is hit within the search radius of R_{th}, the user obtains access to it; otherwise goes to (3);
2) If the downlink SIR_S is larger than the threshold Γ_S, the user is then served by the SCs with the local disk to achieve the throughput of T_S and the procedure ends; otherwise, goes to (7);
3) If the cache in the macro cell is hit, the user obtains access to it; otherwise goes to (5);
4) If the downlink SIR_M is larger than the threshold Γ_M, the user is then served by the macro cell with the local disk to achieve the throughput of T_M^{hit} and the procedure ends; otherwise, goes to (7);
5) The requested content is extracted from the external disk in the core network via the limited backhaul, and sent back to the macro cell.
6) If the downlink SIR_M is larger than the threshold Γ_M, the user obtains access to the macro cell to achieve the throughput of T_M^{miss} and the procedure ends; otherwise, goes to (7);
7) The user experiences outages, and the procedure ends.

As illustrated in Fig. 1, e.g., terminal user equipment UE A initially links to the macro cell in C-plane, and then sends the request for trunk 3 to macro cell. The macro cell receives the request and find the nearest small cell SC3 with the requested content 3. Although SC5 is the nearest SC, UE A prefers to obtain access to SC3 with the assistance of macro cell. When there is no SC available with the requested content within the radius of search region R_{th}, e.g., SC7 fails to serve UE B with the requested trunk 2 and SC6 is beyond the radius of R_{th}, the user is associated to the nearest macro cell.

Based on this procedure, the advantage of the software-defined features of the network can be utilized by considering the cache deployment and the access distance simultaneously. In this way, the users can have higher probability to obtain access to the nearest cell and be served with the local disk, contributing to lower latency and energy consumption.

III. Performance Analysis

In this section, we calculate analytically the probability of hit, coverage probability, throughput and energy consumption of the networks and energy efficiency is derived.

A. Probability of Hit

The cache hit probability of small cell is defined as the probability of existence at least one SC storing the requested content in the coverage area of one user. The coverage radius is the search radius threshold R_{th}. Based on the property of SPPP distribution [8], the probability of hit in macro cell is:

$$P_{hit}^{M} = \sum_{f=1}^{M/L} \frac{f^{-\upsilon}}{\sum_{k=1}^{N_f} k^{-\upsilon}}, \tag{4}$$

The probability of hit in small cell is expressed as follows:

$$P_{hit}^{S} = 1 - e^{-\lambda_S P_c \pi R_{th}^2}, \tag{5}$$

where P_c is the probability of the requested content is in SC:

$$P_c = \frac{S}{S_f} \sum_{f=1}^{S_f/L} \frac{f^{-\upsilon}}{\sum_{k=1}^{N_f} k^{-\upsilon}} \tag{6}$$

In addition, the probability density function (PDF) of the random distance r is derived:

$$f_c(r_{Su}) = 2\pi\lambda_S P_c r_{Su} e^{-\lambda_S P_c \pi {r_{Su}}^2}, \tag{7}$$

which is the probability of the required content exists within R_{th}. The pdf of the nearest distance to the macro cell is:

$$f_r(r_{Mu}) = 2\pi\lambda_M r_{Mu} e^{-\pi\lambda_M {r_{Mu}}^2}, \tag{8}$$

B. Service Coverage Probability

The service coverage provided by the cache-enabled software defined networks can be summarized as follows:

- Obtaining access to SC with the probability of P_{serve}^{S} to achieve the throughput T_S;

$$P_{serve}^{S} = P_{hit}^{S}, \tag{9}$$

- Obtaining access to macro cell and served by the local disk, with the probability of P_{serve}^{M}, to achieve the throughput T_M^{hit};

$$P_{serve}^{M} = (1 - P_{serve}^{S})P_{hit}^{M}, \tag{10}$$

- Obtaining access to macro cell served by the remote delivered content from the backhaul, with the probability of P_{serve}^{B}, resulting the throughput of T_M^{miss}.

$$P_{serve}^{B} = 1 - P_{serve}^{S} - P_{serve}^{M}. \tag{11}$$

The coverage probability is the probability of a random selected user could achieve certain thresholds of SIR Γ_S and Γ_M from small cell or macro cell under the mentioned access strategy in U-plane, which are the indicators of the quality of service. It should be noted, there is no interference between macro cell and SC layer, and the sleep-mode SCs have no interference impact on other SCs in the downlink. With the assistance of macro cell, we assume that the SC can work in semi-sleep mode and transfer the work state immediately. The probability of one SC turns into active mode is that there is at least one active users within the coverage of its Voronoi cell region, which is expressed as follows according to [11]:

$$P_{active} = 1 - (1 + 3.5^{-1}\tfrac{\lambda_u}{\lambda_S})^{-3.5}. \tag{12}$$

The service coverage probability of SC is :

$$\begin{aligned} &P_{Cov}^{S} = E\left[P[SIR_S > \Gamma_S]P_{serve}^{S}\right] \\ &\overset{(a)}{=} \int_0^{R_{th}} \exp\left(-\pi\lambda_S P_{active} r_{Su}^2 \Gamma_S^{\frac{2}{\alpha}} G_\alpha(0)\right) f_c(r_{Su}) dr_{Su} \\ &= \frac{P_c\left(1-\exp\left(-\pi\lambda_S R_{th}^2\left(P_c+P_{active}\Gamma_S^{\frac{2}{\alpha}}G_\alpha(0)\right)\right)\right)}{P_c+P_{active}\Gamma_S^{\frac{2}{\alpha}}G_\alpha(0)}, \end{aligned} \tag{13}$$

where in (a) the conditional joint complementary cumulative distribution function $P[SIR_S > \Gamma_S]$ is given in [12]. The gamma function is defined as $G_\alpha(y) = \int_y^\infty \frac{1}{1+x^{\frac{\alpha}{2}}}dx, y \geq 0$:

$$G_\alpha(y) = \begin{cases} 1/\sin c\left(\frac{2}{\alpha}\right), y- = 0 \\ cot^{-1}y, \alpha = 4, y > 0 \\ \frac{2y_2F_1\left(1,1;2-\frac{2}{\alpha};[1+y^{\frac{\alpha}{2}}]^{-1}\right)}{(\alpha-2)\left(1+y^{\frac{\alpha}{2}}\right)}, y > 0 \end{cases} \tag{14}$$

It should be noted that the interference is accumulated from the distance of zero, because the nearest SC may not always have the requested content in the proposed procedure. The service coverage probability of macro cell with local disk is:

$$\begin{aligned} &P_{Cov}^{M} = E\left[P[SIR_M > \Gamma_M]P_{serve}^{M}\right] \\ &= P_{serve}^{M}\int_0^\infty \exp\left(-x\left(1+\Gamma_M^{\frac{2}{\alpha}}G_\alpha(\Gamma_M^{-\frac{2}{\alpha}})\right)\right)dx \\ &= \frac{P_{serve}^{M}}{1+\Gamma_M^{\frac{2}{\alpha}}G_\alpha(\Gamma_M^{-\frac{2}{\alpha}})}, \end{aligned} \tag{15}$$

where the serving macro cell under the proposed user association strategy is the nearest macro cell.

The service coverage probability of macro cell with the remote delivered content from the backhaul is given herein:

$$\begin{aligned} &P_{Cov}^{B} = E\left[P[SIR_M > \Gamma_M]P_{serve}^{B}\right] \\ &= \frac{P_{serve}^{B}}{1+\Gamma_M^{\frac{2}{\alpha}}G_\alpha(\Gamma_M^{-\frac{2}{\alpha}})}. \end{aligned} \tag{16}$$

The overall coverage probability in the cache-enabled network with software-defined networks is

$$\begin{aligned} &P_{Cov} = P_{Cov}^{S} + P_{Cov}^{M} + P_{Cov}^{B} \\ &= \frac{P_c\left(1-\exp\left(-\pi\lambda_S R_{th}^2\left(P_c+P_{active}\Gamma_S^{\frac{2}{\alpha}}G_\alpha(0)\right)\right)\right)}{P_c+P_{active}\Gamma_S^{\frac{2}{\alpha}}G_\alpha(0)} \\ &+ \frac{e^{-\lambda_S P_c \pi R_{th}^2}}{1+\Gamma_M^{\frac{2}{\alpha}}G_\alpha(\Gamma_M^{-\frac{2}{\alpha}})} \end{aligned} \tag{17}$$

C. Throughput

In the cache-enabled HCNs, the throughput discussed here is the "successful delivery throughput", which is expressed as:

$$\begin{aligned} &T_S = W_S E\left[\log_2\left(1+SIR_S(r_{Su})\right)|SIR_S > \Gamma_S\right]P_{serve}^{S} \\ &= W_S \log_2(1+\Gamma_S)P_{serve}^{S} \\ &+ \frac{W_S}{\ln 2}\int_0^{R_{th}}\int_{\Gamma_S}^\infty \frac{P(SIR_S(r_{Su})>t)}{P(SIR_S(r_{Su})>\Gamma_S)(1+t)}dt f_c(r_{Su})dr_{Su} \\ &\overset{(a)}{=} W_S\log_2(1+\Gamma_S)\left(1-e^{-\lambda_S P_c\pi R_{th}^2}\right) + \frac{W_S P_c}{\ln 2} \\ &\int_{\Gamma_S}^\infty \frac{1-\exp\left(-\pi\lambda_S R_{th}^2\left(P_c+P_{active}G_\alpha(0)\left(t^{\frac{2}{\alpha}}-\Gamma_S^{\frac{2}{\alpha}}\right)\right)\right)}{\left(P_c+P_{active}G_\alpha(0)\left(t^{\frac{2}{\alpha}}-\Gamma_S^{\frac{2}{\alpha}}\right)\right)(1+t)}dt, \end{aligned} \tag{18}$$

where (a) is derived by substituting from (7) (13) into (18). The throughput achieved by macro cell with the local disk is:

$$\begin{aligned} &T_M^{hit} \\ &= W_M E\left[\log_2\left(1+SIR_M(r_{Mu})\right)|SIR_M > \Gamma_M\right]P_{serve}^{M} \\ &= W_M\log_2(1+\Gamma_M)e^{-\lambda_S P_c\pi R_{th}^2}\frac{\sum_{f=1}^{M/L} f^{-\upsilon}}{\sum_{k=1}^{N_f} k^{-\upsilon}} + \frac{W_M}{\ln 2} \\ &\frac{\sum_{f=1}^{M/L} f^{-\upsilon}}{\sum_{k=1}^{N_f} k^{-\upsilon}}\int_{\Gamma_M}^\infty \frac{e^{-\lambda_S P_c\pi R_{th}^2}}{1+t^{\frac{2}{\alpha}}G_\alpha(t^{-\frac{2}{\alpha}})-\Gamma_M^{\frac{2}{\alpha}}G_\alpha(\Gamma_M^{-\frac{2}{\alpha}})}\frac{1}{1+t}dt, \end{aligned} \tag{19}$$

where the overhead cost of the macro cell O_M should be considered. As discussed in [9] [10], the effective macro cell bandwidth in software-defined networks can be modelled as $W_M = (1 - O_M)W$, with $O_M \approx 28.5\%$ and W is the system bandwidth. For small cell, the overhead can be neglected with the assistance of macro cell, so that $W_S \approx W$ can be achieved.

The throughput delivered from macro cell with the remote delivered content from the backhaul is:

$$\begin{aligned} &T_M^{miss} = C(\lambda_M)\cdot P_{Cov}^{B} \\ &= \frac{C(\lambda_M)(1-P_{serve}^{S})(1-P_{hit}^{M})}{1+\Gamma_M^{\frac{2}{\alpha}}G_\alpha(\Gamma_M^{-\frac{2}{\alpha}})}, \end{aligned} \tag{20}$$

where $C(\lambda_M)$ is the capacity limitation of the backhaul per macro cell. It is given by:

$$C(\lambda_M) = \frac{C1}{\lambda_M} + C2, \tag{21}$$

where $C1 \geq 0$ is the backhaul capacity limitation in a macro cell region, and $C2 \geq 0$ is some arbitrary co-efficiency.

The overall network throughput in one macro cell area is:

$$T_{All} = P_{active}\frac{\lambda_S}{\lambda_M}T_S + T_M^{hit} + T_M^{miss}. \tag{22}$$

TABLE I
SIMULATION PARAMETERS.

	Parameter	Value	Parameter	Value
Basic Parameters	α	4	O_M	28.5%[9][10]
	Γ_M, Γ_S	$-10dB$	W	$10MHz$
	λ_S	$0.1/m^2$	λ_M	$0.01\lambda_S$
	R_{th}	$10m$	P_{ow}^{sleep}	$17.2W$[13]
Power Consumption	P_M^t	$46dBm$[13]	P_S^t	$30dBm$[13]
	α_M	3.22[13]	α_S	15.13[13]
	P_M^0	$724.6W$[13]	W_{CA}	$6.25pW/bps$[5]
	P_S^0	$10.16W$[13]	W_{BH}	$50uW/bps$[5]
Cache Parameters	$C1$	$300Mbps$	$C2$	$0Mbps$
	L	$240Mbits$[5]	N_f	1000
	M	$500L$	S_f	$50L-150L$
	S	$5L-50L$	v	0.8[5]

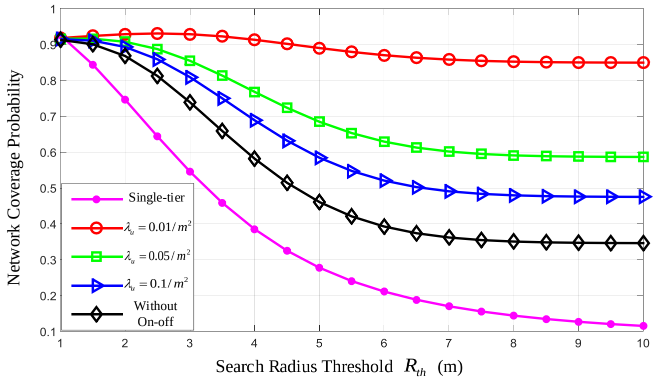

Fig. 3. Network coverage probability in the cache-enabled hyper-cellular networks scaling with the search radius threshold R_{th}: λ_u is the density of user per m^2, with $S_f = 100L$ and $S = 50L$.

D. Energy Consumption

For a typical macro cell area, the overall power consumption is the consumed power by the macro cell and SCs:

$$Pow = Pow_M + \frac{\lambda_S}{\lambda_M} Pow_S, \tag{23}$$

where the power consumption of each SC is:

$$\begin{aligned} Pow_S = & \mathrm{P}_{\text{active}} \left(\alpha_S P_S^t + P_S^0 + w_{CA} T_S \right) \\ & + \left(1 - \mathrm{P}_{\text{active}}\right) P_{ow}^{sleep}, \end{aligned} \tag{24}$$

and the power consumption of a macro cell is:

$$Pow_M = \alpha_M P_M^t + P_M^0 + w_{CA} T_M^{hit} + w_{BH} T_M^{miss}, \tag{25}$$

where α_S and α_M are the increase power co-efficiencies of SC and macro cell, P_S^0 and P_M^0 are the static power consumption of SC and macro cell, P_{ow}^{sleep} is the power consumption of SC in sleep mode, the w_{CA} is the power efficiency of caching hardware in watt per bit of high-speed solid state disk, and w_{BH} is the power consumption per backhaul capacity.

E. Energy Efficiency

The energy efficiency of the network is defined as the ratio of the total network throughput to the energy consumption, with units of bits/Joule, which can be expressed as:

$$EE = \frac{T_{All}}{Pow}. \tag{26}$$

IV. NUMERICAL RESULTS

In this section, we evaluate the impact of deploying cache in the HCNs with the proposed procedure. The main simulation parameters are given in Table I.

In Fig. 3, we provide the analytical results of the network coverage probability versus the search radius threshold R_{th}. Comparisons are also made with the conclusions in single-tier network [8]. With the increase of R_{th} in the proposed access scheme, the UE have higher probability to find a SC cached with the requested contents. So the probability of hit in SCs increases with R_{th}. However, it also brings terrible interference as the nearest SC is not selected as the serving cell. The network coverage probability is influenced by both higher hit probability and higher interference. In our proposed scheme, with the software-defined features, the coverage can be hugely improved, because the users who fails to hit the cache in SCs, can continue access to macro cells. In the worst interference condition, where no SC can be turned off, the network coverage performance still increases by nearly 200% compared with the single tier network. In addition, with the assistance of macro cell, the lower density of active user λ_u, the higher probability to put SCs into sleep mode to lower down the interference. So the network coverage increases with the decrease of active user density. It can be found that the network coverage probability decreases with R_{th} gradually when the active user density is low, and increases initially and then decreases with R_{th} when the active user density is high.

In Fig. 4, we compare our proposed access scheme with the nearest access scheme scaling with the search radius threshold. In the nearest access scheme, the users are always associated to the nearest small cell or the macro cell. Besides, small cells in the nearest access scheme cache the most popular S/L files instead of randomly selecting popular files.

As there is no search radius in the nearest access scheme, it keeps constant with the increase of R_{th}. In our proposed scheme, the larger search radius threshold, the higher probability to obtain access to SCs and achieve the high speed services from the local disk. So the EE is increasing with R_{th} in the proposed system. If R_th is too small, the er search radius threshold, the higher probability to obtain access to SCs and achieve the high speed services from the local disk. So the EE is increasing with R_{th} in the proposed system. If R_{th} is too small, the users can not find a SC cached with the requested content in the region, and obtain access to macro cell with lower SIR compared with nearest access scheme. In addition, the higher density of users, the more small cells are turned on, so that more throughput and energy efficiency can be derived.

Compared with the nearest access scheme, the proposed scheme can increase the network energy efficiency by nearly 27.1% and 6.8% under the active density of $0.01/m^2$ and $0.05/m^2$ respectively when R_{th} is large. This is because that the proposed scheme can take the advantage of more local caches in SC within the coverage, and simultaneously lower down the interference with the on-off strategy in the HCNs. So it can be easily understood that the EE of the proposed scheme is 2% smaller than the nearest access scheme, as the interference has become the most important element.

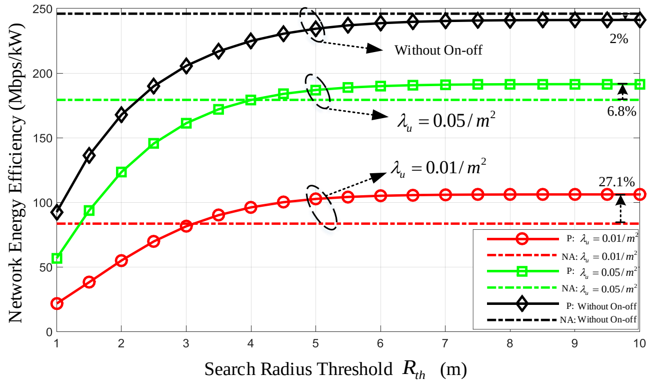

Fig. 4. Network energy efficiency in the cache-enabled hyper-cellular networks scaling with the search radius threshold: λ_u is the density of user per m^2, with $S_f = 100L$ and $S = 50L$. P: proposed access scheme; NA: nearest access scheme.

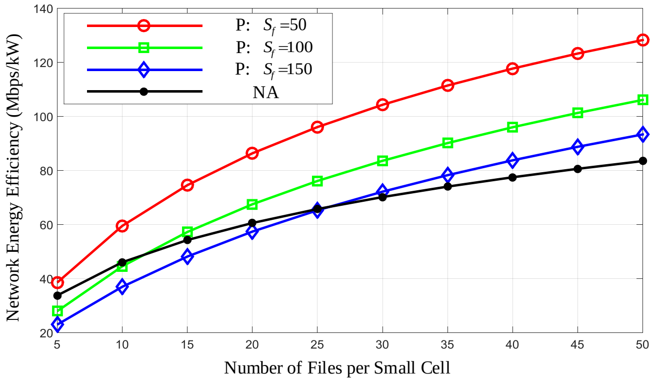

Fig. 5. Network energy efficiency in the cache-enabled hyper-cellular networks scaling with the number of files in small cell: $\lambda_u = 0.01/m^2$. P: proposed access scheme; NA: nearest access scheme.

Fig. 5 gives an illustration on how to improve the performance of the proposed scheme. As expected, the network EE increases with the number of files per SC cached S/L for both proposed scheme and nearest access scheme. However, by shrinking the content library of the small cells S_f, the EE of the proposed scheme can be further improved. It is shown that the EE increases with the decrease of small cell library S_f. By trading off fairness for throughput, the small cells can cache contents more accurately based on user interests. In this way, the files are more easily to be found within the search region. It is illustrated that equipped with small library (e.g., $S_f = 50$), the EE can be improved by more than 54% than the nearest access strategy when active user density $\lambda_u = 0.01$.

It can be concluded that larger search region and cache ability per SC can increase the network EE. Compared with the nearest access scheme, the proposed scheme in HCNs achieves better performance under lower active-user density scenario and can be enhanced further by reducing the SC library.

V. Conclusions and Future Work

In this paper, we investigate how to use cache in the HCNs via taking advantages of group user behaviour and dynamic traffic fluctuation. By proposing a novel access scheme and utilizing on-off strategy in the heterogeneous network, the fairness of service and the user interests are considered. The coverage probability, throughput, energy consumption and EE are derived theoretically and comparisons are made with the single tier network and nearest access scheme. Numerical results show that the HCNs with the proposed scheme can enhance the coverage probability by at least 200% and the EE can be improved by more than 54%. Larger cache ability in SC, smaller content library, lower density of active users and larger search radius can improve the system performance in advance. Future works involves energy harvesting and caching system based on the group user behaviour and energy dynamic changing.

Acknowledgement

This work is supported by the National 973 Program under grant 2012CB316005, by the National Science Foundation of China (NSFC) under grant 61372114 and 61571054, the New Star in Science and Technology of Beijing Municipal Science & Technology Commission (Beijing Nova Program: Z151100000315077) and the China Scholarship Council.

References

[1] L. Jin, Y. Chen, T. Wang, P. Hui and A. V. Vasilakos, “Understanding user behavior in online social networks: a survey,” *IEEE Communications Magazine*, vol. 51, no. 9, pp. 144-150, September 2013.
[2] X. Zhang, Y. Zhang, R. Yu, W. Wang, M. Guizani, “Enhancing spectral-energy efficiency for LTE-advanced heterogeneous networks: a users social pattern perspective,” *IEEE Wireless Communications*, vol. 21, no. 2, pp. 10-17, April 2014.
[3] X. Zhang, R. Yu, Y. Zhang, Y. Gao, M. Im, L. Cuthbert, and W. Wang, “Energy-efficient multimedia transmissions through base station cooperation over heterogeneous cellular networks exploiting user behavior,” *IEEE Wireless Communications*, vol. 21, no. 4, pp. 54-61, August 2014.
[4] K. Wang, Z. Chen, H. Liu, “Push-Based Wireless Converged Networks for Massive Multimedia Content Delivery,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2894-2905, May 2014.
[5] D. Liu, C. Yang, “Will caching at base station improve energy efficiency of downlink transmission?,” *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 173-177, 3-5 Dec. 2014.
[6] C. Yang, Y. Yao, Z. Chen, B. Xia, “Analysis on Cache-Enabled Wireless Heterogeneous Networks,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 131-145, Jan. 2016.
[7] Z. Su, Q. Xu, “Content distribution over content centric mobile social networks in 5G,” *IEEE Communications Magazine*, vol. 53, no. 6, pp. 66-72, June 2015.
[8] S. Tamoor-ul-Hassan, M. Bennis, P. H. J. Nardelli, and M. Latva-aho, “Modeling and analysis of content caching in wireless small cell networks, *CoRR*, vol. abs/1507.00182, 2015. [Online]. Available: http://arxiv.org/abs/1507.00182.
[9] X. Zhang, J. Zhang, W. Wang, Y. Zhang, C.-L. I, Z. Pan, G. Li, Y. Chen, “Macro-assisted data-only carrier for 5G green cellular systems,” *IEEE Communications Magazine*, vol. 53, no. 5, pp. 223-231, May 2015.
[10] J. Zhang, Xing. Zhang, C. Liu, W. Wang, Y. Chen, G. Li, Z. Pan, and C.-L. I, “Theoretical study and performance evaluation of macro-assisted data-only carrier for next generation 5G system”, *International Journal of Communication Systems (IJCS)*, 3rd Feb, 2015.
[11] S. Yu and S-L. Kim, “Downlink capacity and base station density in cellular networks,” *2013 11th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pp. 119-124, 13-17 May 2013.
[12] S. Mukherjee, “Distribution of Downlink SINR in Heterogeneous Cellular Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 575-585, April 2012.
[13] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, A. Fehske, “How much energy is needed to run a wireless network?,” *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40-49, October 2011.