Resource virtualization with edge caching and latency constraint for local B5G operator

Tachporn Sanguanpuak*, Dusit Niyato†, Nandana Rajatheva*, Mehdi Bennis*, Matti Latva-aho*
*6G Flagship, University of Oulu, Finland;
†School of Comp. Science and Eng., Nanyang Technological University (NTU), Singapore
Email: {tachporn.sanguanpuak, nandana.rajatheva, mehdi.bennis, matti.latva-aho}@oulu.fi; dniyato@ntu.edu.sg

Abstract*—The rapidly increasing demand in indoor smallcell networks has given rise to the concept of local beyond 5G (B5G) operator (OP) for local service delivery. The local B5G OP aims to provide wireless network using licensed subbands in an indoor area and tries to gain profits by renting out the infrastructure to the mobile network operators (MNOs). With local B5G OP deployment, the quality of service (QoS) can be guaranteed at mobile broadband users (UEs) and smart devices, i.e., machine type communications (MTC) and ultra reliable low latency (uRLLC). In this paper, we consider the scenario that the local B5G OP aims to maximize profit by optimizing its infrastructure rental fee while renting out cache-enabled smallcell base stations (SBSs) to the MNOs. Each MNO tries to minimize the cache intensity subject to latency constraint at mobile UE. The concept of infrastructure sharing is also deployed at the local B5G OP such that multiple MNOs can utilize the same cache-enabled SBSs simultaneously and the local B5G OP will cache the popular files according to the MNO's largest demand. The optimization problems of the local B5G OP and the MNOs can be transformed into geometric programming problems. Then, we show that the Stackelberg equilibrium is obtained through successive geometric programming (SGP) method. Lastly, we perform an extensive performance evaluation that reveals interesting insights including the optimal SBS intensity that MNOs should rent from the local B5G OP as to satisfy end-to-end latency, 10^{-3} sec, of data transmission from each SBS to UE. The optimal price of renting out infrastructure for the local B5G OP at the Stackelberg equilibrium is also illustrated.

***Index Terms*—local beyond 5G operator (local B5G OP), 6G, edge caching, infrastructure sharing, geometric programming, Stackelberg equilibrium.**

A. Motivation

The 5G and beyond 5G (B5G) technologies will need to support extremely diverse use-cases for example, (i) Extreme mobile broadband (xMBB) with data rates up to several Gbps, more videos, more live streaming and reliable broadband access over large coverage areas. (ii) Massive machine type communications (mMTC) which is a service category consisting of sensing, tagging, and monitoring require high connection density. (iii) Ultra reliable low latency (uRLLC) which is a service category to support the latency-sensitive services such as, remote control, autonomous driving car and tactile Internet [2]. Industrial, manufacturing companies, sport areanas and smart hospitals cannot completely rely on unlicensed wireless band to serve such kind of services. Also, traditional macro cellular networks deployed by the mobile network operators (MNOs) is insufficient to rapidly serve the UEs in an indoor area with the quality of experience (QoE)/quality of service (QoS) garunteeed.

The huge traffic of the above applications is generated from indoor areas [2], the new business model of the MNOs needs to be developed to serve local services with specific requirements [3]–[7]. Therefore, the most prominent and efficient solution is to deploy the concept of local B5G operator (OP) to offer wireless networks for mobile UEs and smart devices in indoor with local licensed spectrum subbands. The facility owner with the capability of deploying small cell base stations (SBSs) in an indoor area with licensed subbands can become local B5G OP [6]–[7]. The local B5G OP can serve the MNO's mobile UEs with licensed subbands while renting out the infrastructure, i.e., SBSs, edge computing servers, and cache storages to the MNOs.

None of the existing work has formally formulated the business model of local B5G OP and multiple MNOs using game theoretical approches while taking latency constraint at each UE into account. The stochastic geometry modeling for BS placement with single/multiple sellers and multiple buyer MNOs and Cournot oligopoly game was proposed in [8]–[9]. Multiple-MNO spectrum sharing using matching game for small cell networks was explored in [10]. Apart from the efficient spectrum and infrastructure utilization, achieving low latency for xMBB with virtual reality services is also a critical challenge for the MNOs. Therefore, the concept of proactive caching was introduced [11]–[15] in which popular contents are stored at the edge/radio access network (RAN), e.g., cache-enabled BSs to reduce the wireless access delay. In the context of economic modeling of caching, the work in [14], the authors considered a Stackelberg game with a single MNO and multiple content providers. The MNO, as the leader, decides on the price to charge to content providers such that the revenue is maximized. The content providers, as the followers, compete with each other to obtain sufficient cache space to improve the QoS to its UEs. In [15], the cache is partitioned into slices and each partition is allocated to the content providers. The utility based approach is used to formulated the problem of content providers.

The majority of the above work consider neither infras-

This paper considers an aspect of the local B5G OP where a full version has been under major revision in IEEE Transaction of Mobile Computing (IEEE TMC) while only an optimization problem of one MNO (MNO's part) has been published in Globecom 2018 [1]

tructure sharing nor the latency constraint at the UE for data transmission in the context of local B5G OP with edge caching. We therefore significantly extend the existing work by developing a framework to model and analyze latency-constrained for radio resource sharing with cache-enabled SBSs in local B5G OP while modeling the location of BSs using stochastic geometry.

The contributions of the paper are as follows:

- We model the local B5G OP that virtualized cache-enabled SBSs to MNOs as a Stackelberg game, where the local B5G OP is the leader and the MNOs are the followers. Each MNO aims to minimize the cost of renting cache intensity from the local B5G OP subject to latency constraint at the mobile UE.
- With the concept of infrastructure sharing, the local B5G OP can deploy cache storage efficiently such that the local B5G OP handles only the largest cache intensity required by the MNOs. The problem of the local B5G OP is formulated as to maximize its revenue while minimizing the power consumption at the BSs. We obtain the Stackelberg equilibrium analytically via successive geometric programming.

B. Organization

The rest of the paper is organized as follows. Section I describes the system model. Section II presents the optimization problem of each MNO as to minimize the *cache intensity* subject to the latency constraint at a UE. This corresponds to the follower subgame in the Stackelberg game formulation. Section III presents the problem of the local B5G OP as to maximize the revenue which corresponds to the leader subgame. The numerical results are presented in Section IV before the paper is concluded in Section V.

I. System Model

We consider a heterogeneous network with a local B5G OP and a set $\mathcal{K}$ of MNOs such that $|\mathcal{K}| = K$. The local B5G OP is assumed to already rented some licensed subbands from multiple MNOs then, attach those licensed subbands to its cache-enabled SBSs while allowing multiple MNOs to simultaneously utilize its cache-enable SBSs. The local B5G OP is assumed to provide the licensed spectrum subbbands while providing a set of cache-enabled SBSs, Φ_b, which are spatially distributed according to a homogeneous Poisson point process (PPP) with spatial intensity λ. Each MNO-k, $k \in \mathcal{K}$, wants to rent a set of cache-enabled SBSs, Φ_k where $\Phi_k \subseteq \Phi_b$ with the SBS intensity, λ_k, from the local B5G OP.

Fig. 1 gives an example for the general case of the cache-enabled BSs required by MNO-1, MNO-2 and MNO-3. Since the local B5G OP deploys resource virtualization, some of the cache-enabled SBSs can be utilized by all three MNOs simultaneously while some are used by one or two MNOs. Each MNO-k operates over orthogonal spectrum, and thus there is no inter-MNO interference. The MNO-k leases out W_k bandwidth with L_k subchannels to the local B5G OP. Each SBS operates in one of the L_k available subchannels

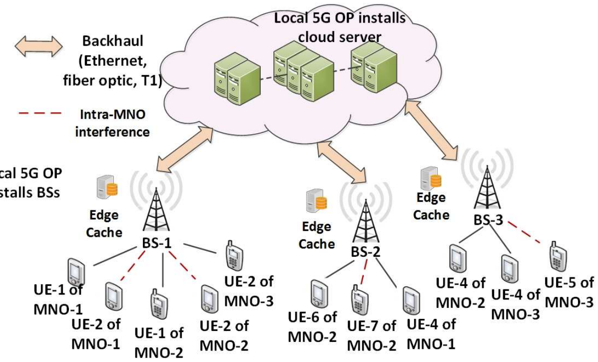

Fig. 1. Virtualized Cache-enabled SBSs shared between MNO-1, MNO-2, and MNO-3.

assigned by the local 5G OP. The subchannel of the same MNO can be accessed by more than one SBS and therefore, the intensity of interfering SBSs which causes intra-MNO interference of the MNO-k is given by $\frac{\lambda_k}{L_k}$. Every BS and UE are assumed to be equipped with a single antenna. For each MNO-k, the SBS serves one UE in a given time slot by using the assigned subchannel with the maximum transmit power p_k. A UE subscribed to an MNO-k associates with the nearest SBS that the MNO-k rents from the local B5G OP. The SBS intensity that a typical UE of the MNO-k can associate itself with is λ_k. First we consider the analysis of a single MNO-k downlink SINR coverage probability, expected data rate and goodput. We assume that the signal undergoes Rayleigh fading with the channel gain, g_k. Let $\alpha_k > 2$ denote the path-loss exponent for the path-loss model $r_k^{-\alpha_k}$, where r_k is the distance between the tagged UE and the nearest SBS-k. Let σ_k^2 denote the noise variance, and again p_k denote the transmit power of each SBS rented by MNO-k. For a given threshold $\bar{T}$, the SINR coverage probability for the tagged UE is defined as $P_c = \mathbb{P}(\mathrm{SINR_k} > \bar{T})$. We use tigh approximation of P_c in Eqn (13) in [8, Proposition 1] but using the intensity of interfering SBSs and the SBS intensity that a typical UE can associate itself with from our case as $\frac{\lambda_k}{L_k}$ and λ_k, respectively. Therefore, the coverage probability can be expressed as,

$$P_c = \left[1 + \frac{\beta - 1}{L_k} + \frac{\alpha}{2\pi\lambda_k\Gamma\left(\frac{2}{\alpha_k}\right)}\left(\frac{\overline{T}\sigma^2}{p_k}\right)^{2/\alpha_k}\right]^{-1}, \quad (1)$$

where $\beta = \frac{2(\overline{T}/p_k)^{2/\alpha_k}}{\alpha_k}\mathbb{E}_{g_k}[g_k^{2/\alpha}(\Gamma(-2/\alpha_k, \overline{T}g_k/p_k)) - \Gamma(-2/\alpha_k)]$ and $\Gamma(z)$ is the Gamma function. For the interference-limited case when $\lambda_k \to \infty$, the last term in (1) will become 0 and the expression of P_c can be simplified to

$$P_c \simeq \frac{L_k}{\beta + L_k - 1}. \quad (2)$$

To derive the expected data rate, given the SINR coverage probability, the downlink transmission rate at a typical UE is,

$$\mathbb{E}[R_k] = \int_0^\infty P_c(e^{\bar{T}} - 1)d\bar{T} = \int_0^\infty P_c(\hat{T})d\hat{T}, \quad (3)$$

where $\hat{T} = e^{\bar{T}} - 1$. The expected rate can be derived using the P_c in (1) for general case as,

$$\mathbb{E}[R_k] = \pi\lambda_k \int_0^{\infty} \Big[\pi\big(\frac{\lambda_k}{L_k}\hat{T}^{2/\alpha}\int_{\hat{T}^{-2/\alpha}}^{\infty}(1+u^{\alpha/2})^{-1}du + \lambda_k\big) + \frac{\alpha}{2}\frac{B^{2/\alpha}}{\Gamma(2/\alpha)}\Big]^{-1} d\hat{T},$$
$$= \pi\lambda_k \int_0^{\infty} \Big[\pi\big(\lambda_I \frac{2\hat{T}}{\alpha-2}\,{}_2F_1\big(1, 1-\frac{2}{\alpha}; 2-\frac{2}{\alpha}; 1-\hat{T}\big) + \lambda_k\big) + \frac{\alpha}{2}\frac{B^{2/\alpha}}{\Gamma(2/\alpha)}\Big]^{-1} d\hat{T}, \quad (4)$$

where ${}_2F_1(a,b,c,z)$ is the Hypergeometric function. We also can derive the goodput of the UE, G_k, for the interference-limited case by using P_c in (2),

$$G_k = \frac{W_k}{\beta + L_k - 1}\log_2(1+\bar{T}). \quad (5)$$

However, in [1], only the goodput, G_k, has been considered.

II. The MNO-k Optimization Problem : Minimization of Cache Intensity with Latency Constraint

In this section, we deal with the minimization of *cache intensity* for each MNO-k where the cache intensity is defined as the product of SBS intensity and cache size. For the interference-limited transmission scenario, we then transform the problem into a geometric program and provide an exact solution.

A. Optimization Problem Formulation

The optimization problem for an MNO-k, where $k \in \mathcal{K}$, so as to minimize the cost of renting cache intensity while satisfying the latency of each typical UE for data transmission is as follows:

$$(P0) \quad \min_{\lambda_k, S_k} \quad \omega\lambda_k S_k \quad (6)$$
$$\text{s.t.} \quad \mathbb{P}(D_k \geq D_{\text{th}}) \leq \epsilon, \quad (7)$$

where $\lambda_k \geq 0$ and $S_k \geq 0$. Here, the latency constraint is given in (7). It is a probabilistic constraint that limits the probability of the delay experienced by a UE while downloading the file, D_k, to be above a certain threshold D_{th} to a small value ϵ, $\epsilon \in (0,1)$. The ω is the price per unit of cache intensity, which is set by the local B5G OP. Since D_k is a random variable whose distribution is not known, in order to make the optimization problem more tractable, we can use the Markov's inequality to linearize the probabilistic constraint in (7). Using Markov's inequality, we have $\mathbb{P}(D_k \geq D_{\text{th}}) \leq \frac{\mathbb{E}[D_k]}{D_{\text{th}}}$. If we ensure that $\frac{\mathbb{E}[D_k]}{D_{\text{th}}} \leq \epsilon$, then the Markov inequality implies that constraint (7) is also satisfied. We show that the delay, D_k is a random variable which depends on λ_k in next subsection.

B. Wireless Access Delay and Hit Probability

In our case, we consider the delay for downloading file only due to the downlink transmission delay, D_{T_x}, if the requested file is already cached at the serving SBS. If the file is not in the cache, the delay experiencd by the UE is the sum of downlink transmission delay plus the backhaul delay as $D_k = D_{T_x} + D_{\text{bh}}$. In order to transfer a file of fixed size x_f, the downlink transmission delay is $D_{\text{Tx}} = \frac{N_k x_f}{G_k}$, where N_k is the number of serving UEs in the cell. The expected number of UEs inside an average Voronoi cell formed by the PPP SBSs is given by $\mathbb{E}[N_k] = \frac{\xi_k}{\lambda_k}$, where ξ_k is the intensity of the UEs.

$$\mathbb{E}[D_{\text{Tx}}] = \frac{\mathbb{E}[N_k]x_f}{G_k} = \frac{\xi_k x_f}{\lambda_k G_k}. \quad (8)$$

We substitute goodput G_k from (5). Since we are using G_k for interference limited case in (5), the G_k of each UE becomes a constant. In (8) the SBS intensity λ_k is a variable that the MNO will need to decide when renting the infrastructure from the local 5G OP. We assume that the SBSs connect to the cloud through optical fiber thus, the backhaul delay, D_{bh}, becomes a constant value. However, the availability of the file in SBS cache is given by the hit probability P_{hit}. The expected total delay is as follows,

$$\mathbb{E}[D_k] = \mathbb{E}[D_{\text{Tk}}]\, P_{\text{hit}} + \mathbb{E}[D_{\text{Tx}} + D_{\text{bh}}]\,(1 - P_{\text{hit}})$$
$$= \mathbb{E}[D_{\text{Tx}}] + \mathbb{E}[D_{\text{bh}}](1 - P_{\text{hit}}). \quad (9)$$

The tight asymptotic approximation of cache hit probability, $P_{\text{hit}}(S_k)$, can be derived as [1, Lemma 2],

$$P_{\text{hit}}(S_k) \simeq \frac{1}{H_{F,\nu}}\left[\zeta(\nu) - \frac{(S_k+1)^{1-\nu}}{\nu - 1}\right], \quad (10)$$

where $\zeta(\nu)$ is the Riemann zeta function. The $\mathcal{F}$ is the set of files available for caching in the clound and $\mathcal{S}_k$ is the set of files can be stored in each SBS of each MNO-k, such that $\mathcal{S}_k \subseteq \mathcal{F}$. We model the popularity of the files by the Zipf distribution and $\nu > 0$ reflects the *skewness* of the file popularity distribution. Then substitute $P_{\text{hit}}(S_k)$ from (10) and $\mathbb{E}[D_{\text{Tk}}]$ from (8) into (9) to obtain the total delay $\mathbb{E}[D_k]$, while D_{bh} is constant value. Finally, the problem (P0) can be transformed into geometric programming problem. The optimal solutions λ_k^* and S_k^* can be computed analytically as in [1, Proposition 4].

III. The Local B5G OP: Model-Based Cache Intensity Pricing for Multiple MNOs

We develop a novel strategy of the local B5G OP for renting out its infrastructure to K MNOs. The SBSs and cache can be shared by multiple MNOs. We assume that the local B5G OP cache sufficient certain amount of files requested by the MNOs. The MNOs request files based on most popular files thus, there is overlapping of the most popular files. For example, let MNO-1 requests $S_1 = 15$ most popular files $\{f_1, \ldots, f_{15}\}$ to be cached and MNO-2 requests $S_2 = 20$ most popular files $\{f_1, \ldots, f_{20}\}$, assuming that the popularity rank of the file corresponds to the file's index, with file f_1 being the most popular and file f_{20} being the least popular. The local B5G OP can satisfy the requests of both MNOs by caching $S_I^* = \max(S_1, S_2) = 20$ most popular files $\{f_1, \ldots, f_{20}\}$, since $\{f_1, \ldots, f_{15}\} \subset \{f_1, \ldots, f_{20}\}$. This is much less than the aggregate amount $S_1 + S_2 = 35$ required to be stored when the cache is not shared among the MNOs.

In general, since $\mathcal{S}_k$ is the set of most popular files requested to be cached by the MNO-k, we can order the sets $\mathcal{S}_k$ as $\mathcal{S}_{\pi(1)} \subseteq \cdots \subseteq \mathcal{S}_{\pi(K)}$, where π represents the permutation of set $\mathcal{K}$. Thus, it is sufficient for local B5G OP to cache the largest set $\mathcal{S}_{\pi(K)}$ of a certain MNO that also meets the demands of all other MNOs. Since the largest set of most popular files also contains the smaller sets of most popular files and the local B5G OP allows cache-enable BSs to be shared among multiple MNOs, the local B5G OP will handle only the largest requested set of files by the MNO denoted as as $\lambda_I^* S_I^* = \max_k\{\lambda_k^* S_k^*\}$. The local B5G OP aims at maximizing its revenue obtained by renting out the cache BSs to the MNOs, while minimizing the power consumption. We assume the maximum interference case from all MNOs, where all K MNOs use the same SBSs simultaneously, as such the transmit power at each SBS will be $p_t = Kp_k + Kp_c$. Since the local B5G OP is renting out λ_I^* SBSs per unit area and, since we assume that all K MNOs use the same SBSs at the same time, the power consumption per unit area is then given by, $Y(\lambda_I^*) = \lambda_I^*(Kp_k + p_c)$, where p_c denotes a fixed amount of circuit power.

The local B5G OP will compute the optimal price of cache intensity, ω^*. The local B5G OP problem is shown in (Q0) in (11). After the local B5G OP computes the price of cache intensity, the local B5G OP will declare the total rent to all MNOs.

A. Optimization Problem of the local B5G OP

When the cache-enable BSs are shared by K MNOs, we can formulate the optimization problem of the local B5G OP as follows:

$$(Q0) \quad \max_\omega \quad \omega\lambda_I S_I - \gamma Y(\lambda_I), \tag{11}$$

where ω is the price of cache per unit area, and γ is the price of areal power consumption, where $\omega, \gamma > 0$. Note that we are not dealing with how other resources, e.g., computing, server, or transmission capacity are shared. We only consider the case where the cache storage in a unit area is shared among the MNOs.

To obtain the solution of the Stackelberg game, we use backward induction method. Accordingly, we first solve the follower subgame problem. This essentially is to solve the optimal strategy of the MNO with the largest required cache intensity. The follower's solution is then used in the leader subgame problem, after which the leader problem is solved. The solution to the leader subgame gives the Stackelberg equilibrium.

Accordingly, we compute the largest cache intensity that the local B5G OP needs to provide to MNOs as $\lambda_I S_I = \max(\lambda_k^* S_k^*)$ by using λ_k^* and S_k^* of each follower MNO-k. We can express λ_k^* and S_k^* in terms of ω as $\lambda_k^* = \frac{T_k}{\omega}$ and $S_k^* = U_k \omega^{-\frac{1}{(\nu-1)}}$, where $T_k = (A+R)/(q_k^*/\omega)$ and $U_k = [V(\nu - 1)/(q_k^*/\omega)]^{1/(\nu-1)}$. Note that, in these expressions for T_k and U_k, from [1, Proposition 4], the term q_k^*/ω is independent of ω, making T_k and U_k independent of ω as well. This transforms the first term of (11), which is $\omega \max_k\{\lambda_k^* S_k^*\} = \omega\lambda_I^* S_I^*$, into

$$\max_k\{U_k T_k\}\, \omega^{-1/(\nu-1)} = UT\omega^{-1/(\nu-1)}.$$

That is, $UT = \max_k\{U_k T_k\}$. Also, the second term in (11) is transformed into

$$\gamma Y(\lambda_I^*) = \lambda_I^* \gamma(Kp_k + p_c) = T\overline{p}\omega^{-1},$$

where $\overline{p} = \gamma(Kp_k)$. Therefore, we can rewrite the maximization problem in (11) as an equivalent minimization problem:

$$(Q1) \quad \min_{\omega>0} \quad T\overline{p}\omega^{-1} - UT\omega^{-1/(\nu-1)}. \tag{12}$$

The problem (Q1) is a signomial optimization problem over the price variable ω. In general, the problem (Q1) is a non-convex problem. However, this problem becomes convex at some values of ν. The solution of the problem (Q1) gives the global optimal solution to the problem (Q0). In order to solve (Q1), let us introduce an auxiliary variable $z \geq 0$ such that it upper bounds the objective function in (12) as follows:

$$z \geq T\overline{p}\omega^{-1} - UT\omega^{-1/(\nu-1)}. \tag{13}$$

Since minimizing the upper bound z minimizes the objective function in (12) as well, the problem (Q1) can be equivalently re-written in terms of this auxiliary variable as follows:

$$(Q2) \quad \chi = \min_{\omega>0} z \tag{14}$$

$$\text{s.t.} \quad \frac{T\overline{p}\omega^{-1}}{z + UT\omega^{-1/(\nu-1)}} \leq 1. \tag{15}$$

Here, the constraint in (15) is obtained after some algebraic manipulations of the bound in (13). To see the equivalence, for fixed ω, the optimal value of $z$is $z = T\overline{p}\omega^{-1} - UT\omega^{-1/(\nu-1)}$. The problem (Q2) is also referred to as a complementary geometric program. We can obtain the solution to the problem in (Q2) via successive geometric programming (SGP) which is equivalent to the solution of (Q0). More details are described in the journal version.

IV. Numerical Results

The transmit power of BS of each MNO-k is $p_k = 1$ Watt, noise power is $\sigma^2 = -150$ dBm, the number of video files in the cloud is $F = 10^5$, the size of the file requested by each UE is $x_f = 10^5$ bits, path-loss exponent is $\alpha = 5$, i.e., suburban area without line of sight [19]. The SINR threshold is $\bar{T} = 10$ dB. Each MNO-k is assumed to have the same number of subchannels as $L_k = 4$. The latency of data transmission should be less than 10^{-3} sec therefore, we limit the total delay to be $\mathbb{P}(D \geq 10^{-3}) \leq 0.01$ in (7). The backhaul delay through optical fiber is assumed to be $\mathbb{E}[D_{\text{bh}}] = 0.0051$ sec.

A. Optimal Price (ω^*) and Maximum Profit (z^*) of the local B5G OP

The local B5G OP will then compute the optimal price ω^* of the infrastructure so as to maximize its profit z^* by using the successive geometric programming algorithm (SGP). In Figs. 2–3, we demonstrate the optimal strategy of the leader local B5G OP at the Stackelberg equilibrium. We assume the bandwidth and the UE intensity to be $[W_1, W_2, W_3]$= [1GHz, 1.5GHz,2GHz] and $[\xi_1, \xi_2, \xi_3] = [10, 15, 20]/(\pi \times 100^2)$, respectively.

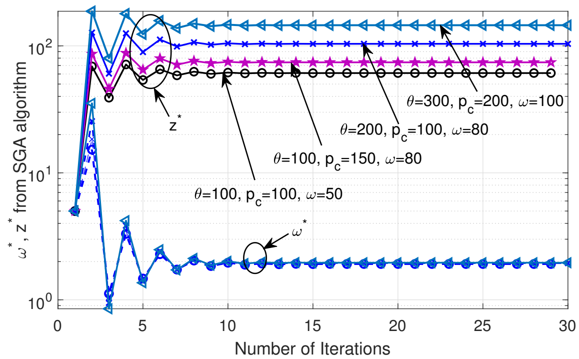

Fig. 2. Convergence of the optimal price of infrastructure (ω^*) and maximum profit (z^*)

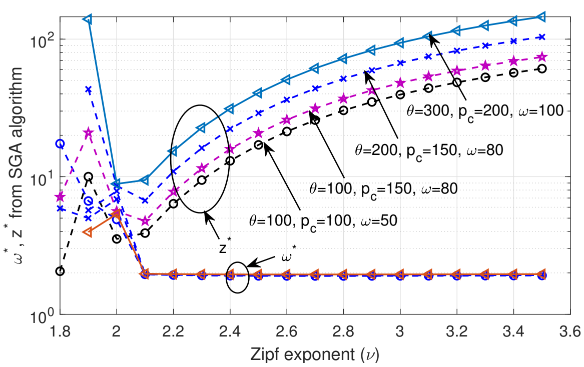

Fig. 3. Optimal price of infrastructure (ω^*) and maximum profit (z^*) versus Zipf exponent (ν).

V. Conclusion

We have proposed a novel deployment of indoor wireless networks for local B5G OP with virtualized cache-enabled BSs. The local B5G OP provides the infrastructure, consisting of SBSs and edge caching, to multiple MNOs. With infrastructure sharing deployment, multiple MNOs are able to use the cache-enabled BSs simultaneously. Each MNO aims to minimize the cost of rented cache intensity subject to latency constraint at each UE while SBSs transmit contents/videos to the UEs. We have modeled the pricing problem for sharing the cache-enabled SBSs infrastructure between the local B5G OP and the MNOs as a Stackelberg game where the local B5G OP is the leader and the MNOs are the followers. Then, we have obtained the optimal strategy of the local B5G OP at the Stackelberg equilibrium via successive geometric programming.

Acknowledgment

This work has been financially supported by the Academy of Finland 6Genesis Flagship (grant no. 318927).

References

[1] T. Sanguanpuak, S. Guruacharya, N. Rajatheva, and M. Latva-aho, "Edge Caching for Cache Intensity under Probabilistic Delay Constraint," *Proc. IEEE Global Commun. Conf.*, pp. 1–7, Dec. 2018.

[2] Cisco, "Cisco visual networking index: forecast and methodology, 2016-2021," *White Paper*, Jun., 2017.

[3] M. Latva-aho, "Micro Operators for Vertical Specifc Service Deliver in 5G," *[Online.] Available: http://5g.ieee.org/images/les/pdf/Workshop8Dec16/D1-Latva-aho-Micro-Operator.pdf*, Jan., 2017.

[4] T. Sanguanpuak, *et al.*, "On spectrum sharing among micro-operators in 5G," *Proc. European Conference on Networks and Communications (EuCNC)*, pp.1–6, 2017.

[5] M.G. Kibria, *et al.*, "Shared Spectrum Access Communications: A Neutral Host Micro Operator Approach," *IEEE Journal on Selected Areas in Commun.*, vol. 35, no, 8, Aug. 2017.

[6] T. Sanguanpuak, D. Niyato, N. Rajatheva, and M. Latva-aho, "Network Slicing with Mobile Edge Computing for Micro-Operator Networks in Beyond 5G," *arXiv:1811.01744*, 2018.

[7] M. Matinmikko, M. Latva-aho, M. Ahokangas, *et al.*,"Micro Operators to Boost Local Service Delivery in 5G," *Wireless Personal Communications* Springer, pp.69-82, Jul. 2017.

[8] T. Sanguanpuak, *et al.*, "Infrastructure sharing for mobile network operators: analysis of trade-offs and market," *IEEE Trans. on Mobile Computing*, 2018.

[9] T. Sanguanpuak, *et al.*, "Inter-operator infrastructure sharing: trade-offs and market," *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, pp.73–78, 2017.

[10] T. Sanguanpuak, *et al.*, "Multi-Operator Spectrum Sharing for Small Cell Networks: A Matching Game Perspective," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3761-3774, 2017.

[11] X. Wang, *et al.*, "Cache in the air: exploiting content Caching and delivery techniques for 5G systems," *IEEE Commun. Magazine*, Feb. 2014.

[12] Z. Chen, *et al.*, "Caching incentive design in wireless D2D networks: a Stackelberg game approach," *Proc. IEEE Int. Conf. on Commun. (ICC)*, Jul. 2016.

[13] J. Li, *et al.*, "Pricing and resource allocation via game theory for a small-cell video caching system," *IEEE Journal Selected Areas in Commun.*, vol. 34, no. 8, Aug. 2016.

[14] F. Shen, *et al.*, "A Stackelberg game for incentive proactive caching mechanisms in wireless networks," *Proc. IEEE Conf. on Global Commun. (GLOBECOM)*, Dec. 2016.

[15] M. Dehghan, W. Chu, P. Nain, and D. Towsley, "Sharing LRU Cache Resources among Content Providers: A Utility-Based Approach,"*arXiv:1702.01823v1 [cs.NI] 6 Feb 2017*.

[16] M. Avriel and A.C. Williams, "An extension of geometric programming with applications in engineering optimization," *Journal of Engineering Mathematics*, vol. 5, no.3, pp 187–194, 1971.

[17] A. J. Morris, "Approximation and complementary geometric programming," *SIAM Journal on Applied Mathematics*, vol. 23, no. 4, pp. 527–531, Dec. 1972.

[18] S.C. Littlechild and G. Owen, "A simple expression for the Shapley value in a special case," *Management Science*, vol. 20, no. 3, Theory Series, 370–372, 1973.

[19] 3GPP TR.36.814 v9.0.0 (2010-03), "3rd Generation partnership project; technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects (Release 9); *http://www.qtc.jp/3GPP/Specs/36814-900.pdf*