Energy Efficient Hybrid Edge Caching Scheme for Tactile Internet in 5G

メタデータ	言語: English 出版者: IEEE 公開日: 2020-11-16 キーワード (Ja): キーワード (En): Energy efficiency, edge computing, hybrid caching, Tactile Internet, 5G 作成者: XU, Jianwen, 太田, 香, 董, 冕雄 メールアドレス: 所属:
URL	http://hdl.handle.net/10258/00010308

Energy Efficient Hybrid Edge Caching Scheme for Tactile Internet in 5G

Jianwen Xu, *Student Member, IEEE,* Kaoru Ota, *Member, IEEE,* and Mianxiong Dong, *Member, IEEE*

***Abstract*—Faster, wider bandwidth and better user experience, 5G is our vision for the future wireless communication. And the Tactile Internet, with ultra low latency, high availability, reliability and security, is going to bring us the unprecedented real-time interactions just like the human sensing. In this paper, we focus on the solving problem of energy efficiency improvement in proactive in-network caching. We design a hybrid edge caching scheme based on four existing methods taking effect in different parts of the network. We also put forward a cache replacement policy to match the hybrid caching scheme considering the popularity of cached files which obeys Zipf distribution. The simulation results show that our proposed methods can reduce latency and achieve better performance in overall energy efficiency than existing ones.**

***Index Terms*—Energy Efficiency, Edge Computing, Hybrid Caching, Tactile Internet, 5G**

I. Introduction

THE next generation of wireless technology is ready for take-off [1], reported by *The Economist* in February, 2018. Since the fourth generation (4G) represented by Long-Term Evolution (LTE) entered the stage of commercial deployment, both industry and academic community have been scrambling to focus on the upcoming fifth generation (5G). And from the user's point of view, everyone is constantly thinking about what innovations the next generation of wireless communication systems will bring to us. Besides faster connection speed, wider bandwidth range and larger throughput, we also consider 5G's performance in terms of user experience and environmental sustainability. The concept of Tactile Internet has emerged as the times require.

Tactile Internet is first defined by the International Telecommunication Union (ITU) as an network architecture that combines ultra low latency with extremely high availability, reliability and security. As the name suggests, Tactile Internet aims to provide a reliable, easy-to-use, low-power, real-time interactive network system just like how human tactile experience be sensed.

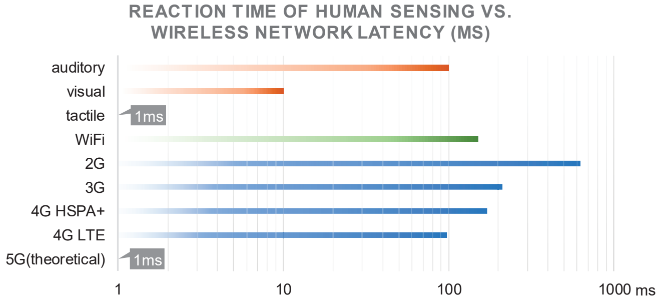


Fig. 1. Reaction time/latency of human sensing and different generations of wireless communications

As shown in Fig. 1, we compare the reaction time/latency of human sensing and different generations of wireless communications. First we choose three traditionally recognized senses, auditory (hearing), visual (seeing) and tactile (touching) as references of how long it takes to sense the outside world through human organs. Red, green and blue bars respectively stand for average reaction time of human sensing [2], ping latency of WiFi and wireless systems from 2G to 4G [3]. From the figure, 4G LTE has reached the comparative latency level of auditory about 100 ms which means we have already achieved the so-called auditory Internet. And for 5G, to reduce the latency of 4G by two orders of magnitude, we skip the stage of visual Internet (10 ms) and set the final goal in Tactile Internet (1 ms).

Initially as an extension of cloud computing at the edge of the network, mobile edge computing (MEC) enables cloud computing capabilities at cellular base stations close by users which can save energy and time cost on the backhaul transmission up to cloud servers. Ever since first raised by European Telecommunications Standards Institute (ETSI) in 2014 [5], we have come to realize that MEC, as a new paradigm, fits well with the urgent needs of large-scale, non-centralized distributed computing in the current Internet of Things (IoT) era. Moreover, MEC is recognized by the European 5G Infrastructure Public Private Partnership (5G PPP) research body as one of the key emerging technologies for 5G networks, with Network Functions Virtualization (NFV) and Software-Defined Networking (SDN) [6].

Proactive in-network caching, which has been receiving great attention in related researches on information centric networking (ICN), is assumed as one of the promising candidate technologies for latency reduction in next generation wireless communication systems [4]. There are mainly four types of caching schemes, local caching, device-to-device (D2D) caching, small base station (SBS) caching and macro base station (MBS) caching.

Fig. 2 gives the four different in-network caching schemes. Local caching refers to caching at the user devices themselves while being requested the same content from the second time [7]. D2D caching bases on D2D communications within

Jianwen Xu, Kaoru Ota and Mianxiong Dong are with the Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan.

Manuscript received xx xx, 2018; revised xx xx, 2019.

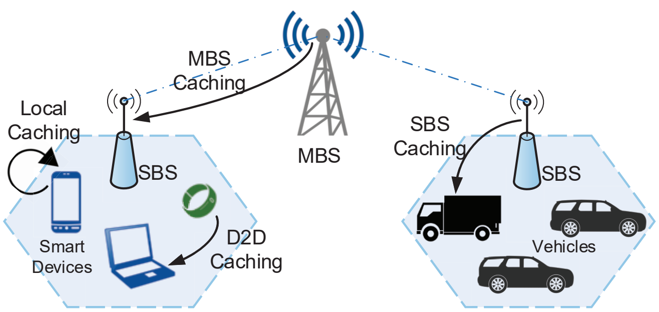


Fig. 2. Different types of proactive in-network caching schemes

small cells [9]. SBS caching and MBS caching are available in respective base stations [10]. Each scheme has its own technical details and almost no overlap of four schemes exists in the actual application scenarios. As a result, in this paper, we try to propose a hybrid edge caching scheme to reduce latency and optimize the overall energy efficiency of edge caching. Our target is to integrate the four existing schemes taking effect in different parts of the network for performance improvement and pave the way for Tactile Internet in the near future.

The main contributions of our work are as follows.

- Design a 3-tier heterogeneous network model including Server Tier, MBS Tier and SBS Tier from top to bottom;
- Propose a cache replacement policy considering the popularity of cached files which obeys Zipf distribution;
- Propose a hybrid edge caching scheme based on existing methods to reduce latency and improve overall energy efficiency;
- The simulation results prove that our proposed methods can optimize the network performance compared with existing ones.

This paper is divided by six sections to elaborate our work on energy efficient edge caching for Tactile Internet. Section 2 introduces related works in the fields of Tactile Internet and hybrid caching. Section 3 describes the mathematical model of a 3-tier heterogeneous network and formulates the problems to solve. Section 4 proposes the cache replacement policy and hybrid edge caching scheme. Section 5 carries out experimental simulations and analyzes the performance of proposed methods. Section 6 summarizes our work.

II. Related Work

In this section, we present related work about 5G and Tactile Internet, mobile edge computing and hybrid caching.

A. 5G and Tactile Internet

Today, with the emergency of information explosion and the IoT boom, 5G is playing an essential role in leading a new round of technological innovation. Boccardi *et al.* summarize five disruptive research directions of 5G, device centric architectures, millimeter wave (mmWave), massive multiple-input multiple-output (MIMO), smarter devices and native support for machine-to-machine (M2M) communication. They analyze both architectural and component design changes of each direction and consider that the corresponding solutions will form the basis of 5G [7]. Shafi *et al.* discuss the standardization trials and deployment challenges of the user centric 5G systems [8].

5G and Tactile Internet can be regarded as the relationship between the technology and its actual deployment. Simsek *et al.* research on the relationship between 5G and Tactile Internet and put forward the concept of 5G-enabled Tactile Internet. They summarize the technical issues and challenges in the cross-field such as end-to-end architecture, hardware and PHY layer transmission approaches, unprecedented edge-cloud and AI capabilities [2]. Maier *et al.* elaborate the commonalities and subtle differences between the Tactile Internet and IoT and 5G. The common features are ultra low latency and high reliability, the co-existence of human-to-human and machine-to-machine communications, data centric technologies and security issues [11]. Ajiaz *et al.* focus on the essential design challenges on achieving haptic communications with 5G cellular networks [12].

Today, 5G has entered a stage of rapid development. Many countries and giant companies are competing with one another to take early actions and seize the business opportunities, telecommunications operators are making great efforts to announce 5G standards and network launches. Zhang *et al.* propose mobility management schemes to achieve seamless handover in network slicing 5G systems [13]. Parvez *et al.* survey all the emerging technologies for satisfying the technical requirement of 5G communication. They consider radio access network (RAN), core network and caching as three main solution domains [4]. Antonakoglou *et al.* pay attention to communication protocols and frameworks, control system approaches needed by haptic communications and 5G [14].

B. Mobile Edge Computing and hybrid Caching

As one of the three key technologies in 5G, MEC aims at providing decentralized solutions to ease the workload of the central cloud server and improve the overall network reliability. Ahmed *et al.* summarize the opportunities, research challenges and possible solutions in facing delay sensitive service demands in accessing cloud center [15]. Tao *et al.* research on the 5G enabled vehicle-to-grid (V2G) networks and design a fog-cloud cooperative computing model [16]. Li *et al.* apply MEC into deep learning and network function virtualization [17] [18]. Mao *et al.* focus on the problem of joint radio and computational resource management, and discuss the key research problems and preliminary solutions [19]. Ford *et al.* design a distributed mobile edge cloud to reduce latency and increase resilience for 5G [20]. Li *et al.* combine MEC with deep learning and IoT, and put forward a offloading strategy to optimize learning performance in IoT applications [21].

With the development of related researches on information-centric networking in recent years, in-network content caching is attracting more and more attention. Caching can further reduce extra transmission consumption by enhancing the reuse of data and contents. A reasonable caching scheme is able to improve 5G system stability without affecting overall network performance. Wang *et al.* learn from caching technologies in current mobile networks and compare the access delay and traffic load of evolved packet core (EPC) caching, RAN caching and ICN-based methods [22]. Bastug *et al.* propose a proactive edge caching mechanism to work at off-peak periods according to popularity and patterns of files, and correlations among users [23]. Li *et al.* present an edge centric computing CCN orchestrating scheme called ECCN for 5G RAN structure [24]. Erol-Kantarci *et al.* apply mobile edge caching into augmented reality (AR) and virtual reality (VR), and overcome the challenges on link capacity and backhaul traffic [25]. Li *et al.* consider a prefetchable memory object caching design for real-time data processing in IoT system [26].

Hybrid caching refers to the methods that take advantage of distributed massive devices can be used for content edge caching and exploit the diversity of caching itself. Tran *et al.* propose a real-time, context aware collaborative caching framework within mobile RAN and prove its promising benefits in facilitating the development of 5G technologies [27]. Kwak *et al.* research on hybrid edge caching in cloud units and base stations to support high content request rate. They focus on the cache control problem to maximize the throughput under the limitation of service latency [28] [29]. Deng *et al.* propose hybrid caching strategies on D2D communication based on Gauss Poisson process to optimize cache hit rate and offloading gain [30].

III. Problem Formulation

In this section, we design a 3-tier heterogeneous Tactile Internet network model and formulate the two problems to solve.

A. System Model

To deploy hybrid edge caching in a Tactile Internet network architecture, and achieve the target of reducing end-to-end latency and improving energy efficiency, first we design a 3-tier network model.

As shown in Fig. 3, from top to bottom there are Server Tier, MBS Tier and SBS Tier. Server Tier is made up of central cloud servers and storage units. As the top tier, Server Tier has the absolute advantage in processing and storage capabilities. Besides, Server Tier owns a complete content library including all original files is the only tier that can handle user requests without regard to in-network caching. MBS Tier consists of macrocells served by cell sites with high signal transmission power. We use tower symbols to represent macro base stations. The light red oval area is used only to identify the current tier which means that the relative positions of MBSs in the figure does not equal to the actual distances. For SBS Tier, we use multiple hexagons as small cells. Compared with MBSs, SBSs provide low-powered cellular radio access. As mobile users (smart devices, vehicles), when we move across the signal scope of small cells, our requests will then be answered by the current SBSs. Both MBS Tier and SBS Tier can provide edge computing service to mobile users.

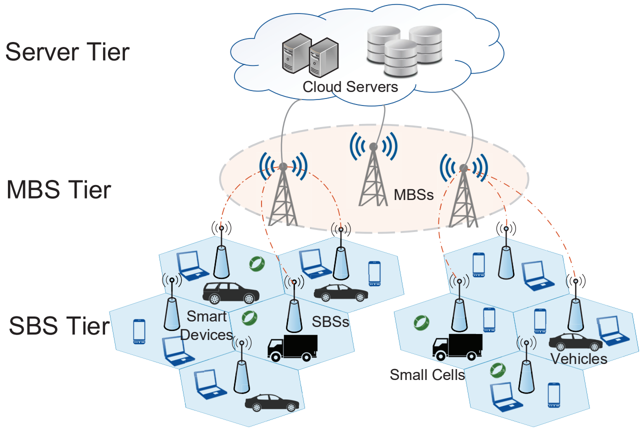


Fig. 3. A 3-tier heterogeneous Tactile Internet network model

In this model, $M = \{m_1, m_2, ...\}$ and $S = \{s_1, s_2, ...\}$, m and s respectively stands for single MBS node and SBS node. Mobile users are $U = \{u_1, u_2, ...\}$ who send out requests for files within content library. Let content library and files in it be denoted by $CL = \{cl_1, cl_2, ..., cl_{n_{CL}}\}$ which are sorted by popularity. n_{CL} is the number of files in CL. As a result, the popularity of one of the n_{CL} files can be expressed as [31]

$$p_{cl_k}(k, x, n_{CL}) = \frac{1/k^x}{\sum_{i=1}^{n_{CL}}(1/i^x)} \tag{1}$$

where $k \in \{1, 2, ..., n_{CL}\}$ stands for the rank or we can say it is the kth popular file in CL. x is value of the exponent which characterizes the Zipf distribution. Zipf distribution or Zipf's law is an empirical law that states the relation between frequency of occurrence of an event and its rank by this frequency with all events [32]. The idea of Zipf's law has long affected the design in the Internet such as content delivery networks and peer-to-peer networks [33]. Here we use it in describing the popularity of files in content library which can help arrange the limited storage capacity of each tier when designing the caching method. That is, we can obtain the popularity (frequency of occurrence) of a given file in CL only from its rank (k) by Equation (1).

B. In-network Content Caching

In the scenario of Fig. 3, mobiles users in the small cells will check if their requests can be answered by cache instead of central server.

When u_i sends a request for cl_k, firstly u_i will check if cl_k already be cached locally. Then within the current small cell, other users may also check if they have it. In the case that both local caching and D2D caching can not come in handy, the current SBSs and MBS will take turns to search cl_k in respective storage units. For local caching and MBS caching, we only consider one edge node, the current user device itself and macro base station. D2D caching needs to traverse all neighbor devices as edge nodes. And for the case of SBS caching, we may get help from all the SBSs within the same MBS.

To calculate the probability that any u_i can get cl_k from cache, we need to consider the four caching methods separately. Here we use $EN = \{en_1, en_2, ...\}$ to represent edge nodes in one of the four methods. That is, the EN can be one of the {local, D2D, SBS, MBS} which means edge nodes here include user device itself (local), user devices nearby (D2D), SBSs and MBSs. A_{EN} is the total coverage area of EN and C_{EN} is the total storage capacity for caching (or how many files can be cached at most), respectively. Assume the number of occurrences that cl_k is not cached in EN within A_{EN} follows a homogeneous Poisson point process [34] [35], we have the cumulative distribution function (CDF) that cl_k is not cached in any en

$$F_{cl_k}^{EN}[no\_cache] = e^{-\lambda_k A_{EN}} \tag{2}$$

where λ_k is the intensity or arrival rate in homogeneous Poisson point process. Here λ_k stands for the expected number of edge nodes that have the cl_k per unit area which means $\lambda_k = \rho_{EN} \cdot P_k$. P_k is the probability that cl_k is cached in EN, and ρ_{EN} is the spatial density of EN in A_{EN}. As a result, the probability that cl_k is cached in at least one en should be

$$F_{cl_k}^{EN} = 1 - F_{cl_k}^{EN}[no\_cache] = 1 - e^{-\rho_{EN} P_k A_{EN}} \tag{3}$$

With Equation (1), the total probability that u_i can be satisfied by edge caching is

$$\begin{aligned} F^{EN} &= \sum_{k=1}^{n_{CL}} p_{cl_k}(k,x,n) \cdot F_{cl_k}^{EN} \\ &= \sum_{k=1}^{n_{CL}} \frac{k^{-x}(1 - e^{-\rho_{EN} P_k A_{EN}})}{\sum_{i=1}^{n} i^{-x}} \end{aligned} \tag{4}$$

n_{CL} is number of files in CL. Thus, the probabilities of four different caching methods are

$$F^{local} = \sum_{k=1}^{n_{CL}} \frac{k^{-x}(1 - e^{-P_k^U})}{\sum_{i=1}^{n} i^{-x}} \tag{5}$$

where P_k^U is the probability that cl_k is cached in U. Since we can use $\rho_{EN} \cdot A_{EN}$ to express the number of en within the coverage area, in the case of local caching, the value of $\rho_{EN} \cdot A_{EN}$ becomes 1.

$$F^{D2D} = \sum_{k=1}^{n_{CL}} \frac{k^{-x}(1 - e^{-\rho_U P_k^U A_{D2D}})}{\sum_{i=1}^{n} i^{-x}} = \sum_{k=1}^{n_{CL}} \frac{k^{-x}(1 - e^{-\rho_U P_k^U \pi r_U^2})}{\sum_{i=1}^{n} i^{-x}} \tag{6}$$

where ρ_U is the spatial density of U, A_{D2D} is the signal coverage area of D2D communications and r_U is the coverage radius of a single user device, respectively.

$$\begin{aligned} F^{SBS} &= \sum_{k=1}^{n_{CL}} \frac{k^{-x}(1 - e^{-\rho_{SBS} P_k^{SBS} A_{MBS}})}{\sum_{i=1}^{n} i^{-x}} \\ &= \sum_{k=1}^{n_{CL}} \frac{k^{-x}(1 - e^{-\rho_{SBS} P_k^{SBS} \pi r_{MBS}^2})}{\sum_{i=1}^{n} i^{-x}} \end{aligned} \tag{7}$$

where ρ_{SBS} is spatial density of S. P_k^{SBS} is probability that cl_k is cached in S. As shown in Fig. 3, each user can have one MBS and multiple SBSs for service. After an SBS receiving the original request from user, it will forward to other SBSs within the same MBS before sending upwards. That is, A_{EN} here equals to the coverage area of multiple SBSs under the same MBS. Or we may say that a user can communicate with all the SBSs under one MBS, but some of them undertake more forwarding hops.

$$F^{MBS} = \sum_{k=1}^{n_{CL}} \frac{k^{-x}(1 - e^{P_k^{MBS}})}{\sum_{i=1}^{n_{CL}} i^{-x}} \tag{8}$$

where P_k^{MBS} is the probability that cl_k is cached in M. Similar with local caching, since $A_{EN} \cdot \rho_{EN}$ can express the number of en within the coverage area, for MBS caching, there is only one MBS to provide service, then the value of $\rho_{MBS} \cdot A_{MBS}$ becomes 1.

C. Problem Formulation

In this paper, we put forward a hybrid edge caching scheme to solve the problem in reducing end-to-end latency and improving energy efficiency. The two main metrics we choose for performance evaluation are end-to-end latency and energy efficiency. End-to-end latency includes the time cost on two-way transmission from user to cloud server. And once the user request can be satisfied by one of the four caching methods, the corresponding part can be saved. Energy efficiency is calculated by the total size of the requested files divided by energy consumption.

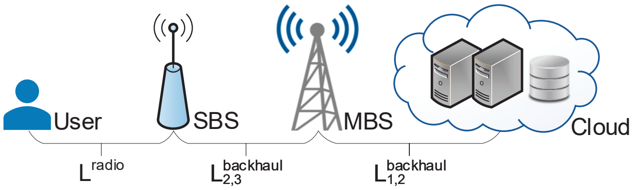


Fig. 4. End-to-end latency

$$L^{e2e} = 2 \times (L^{radio} + L_{2,3}^{backhaul} + L_{1,2}^{backhaul}) \tag{9}$$

L^{e2e} stands for the end-to-end latency. When no cache is found for user request under the current network conditions, the results are shown in Fig. 4 and Equation (9). L^{radio} and $L^{backhaul}$ respectively stands for transmission time between users/edge nodes (SBSs and MBSs), and edge nodes/central cloud servers. For L^{radio}, the communications is wireless broadcast. The case of $L_{2,3}^{backhaul}$ between SBSs and MBSs is wireless point-to-point. The case of $L_{1,2}^{backhaul}$ is wired point-to-point. With edge caching, actually we can save part of the latency on backhaul or routing. For example, the latency of MBS caching L_{e2e}^{MBS} is $2 \times (L^{radio} + L_{2,3}^{backhaul})$.

As a result, the problem of reducing total latency in our 3-tier network model is

$$\begin{aligned} minimum \quad & \mathcal{E}[F^{EN'}L^{EN'}] \\ subject\ to \quad & \sum_{k=1}^{n_{CL}} b_k^{EN}|cl_k| \le C_j^{EN} \\ & \forall j \in \{1,2,...,n_{EN}\} \\ & b_k^{EN} \in \{0,1\} \end{aligned} \tag{10}$$

where $F^{EN'}$ and $L^{EN'}$ are probability and end-to-end latency in answering one user request. The apostrophe after EN means that the case of requests not being answered by cache is also considered in the calculation of latency. $\mathcal{E}$ stands for the expected value. Here we need the results of end-to-end latency on the repetitions of user requests being satisfied by any caching method. To pursue the minimum, we have to make sure that the total size of cached files not exceeds the capacity C_j^{EN} in any edge node. The absolute value notation here means the size of cl_k in content library. We add a boolean variable b_k^{EN} to denote that cl_k is cached in en_j, or not.

Energy cost on edge caching is another aspect we consider in performance evaluation. The longer a user request travels, the more it may spend. Furthermore, the parts of forwarding hops in the routing procedures need extra consumption. As a result, energy consumed in end-to-end transmission during L^{e2e} can be expressed as

$$E^{e2e} = L^{e2e}(p^{tran}/\eta + p^{stat}) \tag{11}$$

E^{e2e} is the energy consumption of end-to-end transmission. p^{tran} and p^{stat} respectively represents transmission power and static power of all other circuit blocks in edge nodes as transmitter and receiver [36] [37]. η is the efficiency of transmit power amplifier.

Then we can calculate total energy efficiency EE [39] as

$$EE = \frac{R \cdot L_{total}^{e2e}}{\sum_{r=1}^{n_t} F^{EN'} E_r^{e2e}} = \frac{R \cdot \sum_{r=1}^{n_t} L_r^{e2e}}{\sum_{r=1}^{n_t} F^{EN'} L_r^{e2e}(p_r^{tran}/\eta + p^{stat})} \tag{12}$$

where R stands for the data rate of the overall network. n_t stands for the number of user requests. Thus, the product of R and L_{total}^{e2e} is the total size of data generated in edge caching.

As a result, the second problem of improving energy efficiency in our 3-tier network model is

$$\begin{aligned} maximum \quad & EE \\ subject\ to \quad & \sum_{k=1}^{n_{CL}} b_k^{EN}|cl_k| \le C_j^{EN} \\ & \forall j \in \{1,2,...,n_{EN}\} \\ & b_k^{EN} \in \{0,1\} \end{aligned} \tag{13}$$

where EE is obtained by Equation (12). Limitation factors are the same with Equation (10).

Besides end-to-end latency and energy efficiency, we also take into consideration cache hit ratio. cache hit ratio here plays the role of showing how caching schemes can maximize the number of cache hits and minimize the number of misses.

$$Cache\ hit\ ratio = \frac{Cache\ hits}{Cache\ hits + Cache\ misses} \tag{14}$$

As shown in Equation (14), to calculate cache hit ratio, we need to respectively count the times that cache hits and misses within a given period of time. High cache hit ratio can show the probability that the user request is satisfied by cache.

TABLE I summarizes and lists the main symbols used in this paper.

TABLE I
Notations in hybrid edge caching for Tactile Internet

Symbol	Meaning
M, m	set of MBSs in MBS Tier and one in it
S, s	set of SBSs in MBS Tier and one in it
U, u	set of mobile users and one in it
CL, cl	set of files in content library and one in it
p_{cl}	the popularity of cl
EN, en	set of edge nodes and one in it
A_{EN}	total coverage area of EN
C_{EN}	total cache capacity of EN
F_{cl}^{EN}	the probability that u_i can receive response from cache in EN
λ	intensity of homogeneous Poisson point process
ρ_{EN}	spatial density of EN in A_{EN}
P_{EN}	the probability that cl_k is cached in EN
F^{EN}	the total probability that u_i can receive response from cache in EN
r_{EN}	coverage radius of en in EN
$L^{e2e}, L^{radio}, L^{backhaul}$	latency of end-to-end, transmission between users and edge nodes and backhaul to Server Tier
n_t	number of request times from U
b^{EN}	boolean variable to denote that cl is cached in any en or not
E^{e2e}, EE	energy consumption of end-to-end transmission and energy efficiency
t_{tran}	duration time of transmission
p^{tran}, p^{stat}	transmission power and static power of transmitter and receiver
η	efficiency of transmit power amplifier
R	date rate of the overall network

IV. hybrid Edge Caching Scheme for Tactile Internet

In this section, we propose a hybrid edge caching scheme to solve the two problems and help achieve Tactile Internet in low latency and high energy efficiency.

To reduce the total latency and improve energy efficiency of edge caching, we integrate four existing schemes into a hybrid one which considers all the computing resources available within the network. Moreover, since the capacities of edge devices are limited, we have to discard parts of the cached data or files when memories are full. We need a suitable cache replacement policy to help each cached file do its best in saving as much time as possible on backhaul transmissions. Our target is to ensure that files with high popularity can extend TTL by migrating between cache segements instead of coming out. As a result, when we reasonably occupy the cache capacity of each edge device, and improve the utilization of files during their cache TTL as a whole, time cost required for upward transmission will be greatly reduced. At the same time, energy efficiency given by Equation (12) and (13), which is calculated as the total size of the requested files devided by total energy consumption. And energy consumption (Equation (11)) is also calculed from time cost. Therefore, when we cut down the latency, energy efficiency can be improved correspondingly.

A. Cache Replacement Policy

In the design of caching scheme, cache replacement policies refers to the algorithms choosing to keep or discard the cached items when cache capacity is full. As a result, each item in cache may own a time-to-live (TTL) which indicates the length of time it can stay in the cache memory. Actually, different cache replacement policies use different rules to define and calculate TTL. As one of the most common policies, first in first out (FIFO) maintains a FIFO queue and evicts the items entering first when any newcomer has no position. Least recently used (LRU) monitors all the cached items and evicts the one used least recently. Random replacement (RR) chooses item to evict without considering any information about history or forecast. RR is often used as a benchmark for performance comparisons, but it can also have unexpected results in some cases. That is, no cache replacement policy is always the best. And what we want is to find the most suitable one for our caching scheme.

In the scenario of solving the two problems in the proposed 3-tier network model, to minimize end-to-end latency and maximize energy efficiency we need a cache replacement policy to distinguish the popularity of the files in CL and distribute limited capacities of edge devices to the most popular ones that may be requested frequently.

As shown in Algorithm 1, req is the requested file from user and $evict$ is the file to be evicted from cache. Q_{seg1} and Q_{seg2} respectively stands for the queues of cached files in two segments with different priorities. $left$ means the rest of the cache capacity. $size$ is the size of request file. $pop()$ and $push()$ are the pop and push functions. $head$ is the cached file at the queue head. $any()$ is the function telling whether a file is cached in the queue. Here we propose a double segmented LRU cache replacement (S2LRU) policy for hybrid edge caching for Tactile Internet. The motivation of Algorithm 1 is to design a optimized cache replacement approach suitable for hybrid edge caching. Key idea of Algorithm 1 is that with more than one cache segments, files of different popularity are treating separately. Popular files will not be easily discarded, and relatively unpopular files will be replaced soon. We divide the total cache capacity into two prioritized parts. The Q_{seg1} with low priority is checked first, once the cache hits, the cached file will leave Q_{seg1} and move to higher Q_{seg2}. Then we check the Q_{seg2}, once the cache hits, the cached file come to the queue's tail. And for files cached in Q_{seg2}, once the left capacity ($C_{seg2}.left$) is not enough to accommodate any newcomer, files at the queue's head ($Q_{seg2}.head$) may have to walk out of Q_{seg2} and come back to Q_{seg1}. When the same situation occurs in Q_{seg1}, files will be evicted from the current en.

Algorithm 1 Segmented LRU for Hybrid Edge Caching

```
$req, evict$ ←requested file from user and evicted file from cache
$Q_{seg1}, Q_{seg2}$ ←queues of two segmented cache capacity
if $any(Q_{seg1} = req)$ then
    $C_{seg1}.left \leftarrow C_{seg1}.left + req.size$
    $Q_{seg1}.pop(req)$
    while $C_{seg2}.left < req.size$ do
        $evict_{seg2} \leftarrow Q_{seg2}.head$
        $C_{seg2}.left \leftarrow C_{seg2}.left + evict_{seg2}.size$
        $Q_{seg2}.pop()$
        while $C_{seg1}.left < evict_{seg2}.size$ do
            $C_{seg1}.left \leftarrow C_{seg1}.left + Q_{seg1}.head.size$
            $Q_{seg1}.pop()$
        end while
        if $C_{seg1}.left \geq evict_{seg2}.size$ && $!any(Q_{seg1} = evict_{seg2})$ then
            $Q_{seg1}.push(evict_{seg2})$
            $C_{seg1}.left \leftarrow C_{seg1}.left - evict_{seg2}.size$
        end if
    end while
    if $C_{seg2}.left \geq req.size$ && $!any(Q_{seg2} = req)$ then
        $Q_{seg2}.push(req)$
        $C_{seg2}.left \leftarrow C_{seg2}.left - req.size$
    end if
else if $any(Q_{seg2} = req)$ then
    $Q_{seg2}.pop(req)$
    $Q_{seg2}.push(req)$
else
    while $C_{seg1}.left < req.size$ do
        $Q_{seg1}.pop()$
    end while
    $Q_{seg1}.push(req)$
end if
```

The proposed S2LRU takes advantage of several common policies. First, as shown in the title, we extend the TTL of cached files that are recently used and correspondingly make the files rarely used easy to expire. Second, least-frequently used (LFU) policy also plays a role in S2LRU since only files being requested more than one time may enter Q_{seg2}. Third, we choose double segmented LRU rather than triple (S3LRU) or higher because of the limited capacity of edge devices. Multiple segments may lead to fragmentation of the cache space, making the efficiency drop dramatically when processing large files. Time complexity of Algorithm 1 is $O(n_{CL})$ which stands for the worst case that all the files in CL are cached in this en and the requested one is found lastly.

B. Hybrid Edge Caching Scheme

In this subsection, we design a hybrid edge caching scheme based on the four types of methods shown in Fig. 2. Our target is to coordinate all the computing resources available in network model in Fig. 3, and reasonably arrange the caching and discarding of files with different popularities under the Zipf distribution.

Algorithm 2 HECT: Hybrid Edge Caching for Tactile Internet

```
$u, u.in\_sbs, u.in\_sbs.in\_mbs$ ←current user, the SBS and MBS beyond current user
$n_{d2d}$ ←number of other users within the signal scope of $u$
$n_{sbs}$ ←number of SBSs within the same MBS
 1: if $find(u.cache = req)$ then
 2:     calculate $L$ and $EE$ by local caching
 3: else
 4:     for $i \leftarrow 1$ to $n_{d2d}$ do
 5:         if $find(u_i^{d2d}.cache = req)$ then
 6:             calculate $L$ and $EE$ by D2D caching
 7:             cache $req$ at $u$
 8:         end if
 9:     end for
10:     if $find(u.in\_sbs.cache = req)$ then
11:         calculate $L$ and $EE$ by SBS caching
12:         cache $req$ at $u$
13:     else
14:         for $j \leftarrow 1$ to $n_{sbs}$ // routing among all SBSs under the same MBS do
15:             if $find(sbs_j^{u.in\_sbs.in\_mbs}.cache = req)$ then
16:                 calculate $L$ and $EE$ by SBS caching
17:                 cache $req$ at $u$ and all $sbs$ in the forwarding route
18:             end if
19:         end for
20:         if $find(u.in\_sbs.in\_mbs.cache = req)$ then
21:             calculate $L$ and $EE$ by MBS caching
22:             cache $req$ at $u$ and $u.in\_sbs$
23:         else
24:             calculate $L$ and $EE$ by no caching
25:             cache $req$ at $u$, $u.in\_sbs$ and $u.in\_sbs.in\_mbs$
26:         end if
27:     end if
28: end if
```

As shown in Algorithm 2, u is the current user sending out request. $u.in\_sbs$ and $u.in\_sbs.in\_mbs$ are the SBS first receiving the user request and the MBS. n_{d2d} is the number of other users that can communicate with the current user in D2D mode. n_{sbs} is the number of SBSs within the signal of the same MBS. $find()$ is the function telling if the requested file can be found in the cache, TRUE for found and FALSE for not. The motivation of Algorithm 2 is to set rules for caching on any of the four parts based on the replacement policy in Algorithm 1. The key idea of Algorithm 2 is that from a global perspective, making full use of each device in the network architecture that can provide cache storage.

TABLE II
EXPERIMENTAL SETUPS

Parameter	Value
Simulation area	$950\sqrt{3} \times 1300\ m^2$
Number of files in CL	200
Number of users/SBSs/MBSs	1000/100/4
Cache capacity of users/SBSs/MBSs	$2/10/50\ MB$
Signal scope radius of D2D/SBS	$20/100\ m$
Date rate in 5G	$10\ Gbps$
Wave propagation speed	**c** (speed of light)
User power in D2D communication	$100\ mW$
SBS/MBS power	$2/100\ W$
Distance between adjacent SBSs	$100\sqrt{3}\ m$

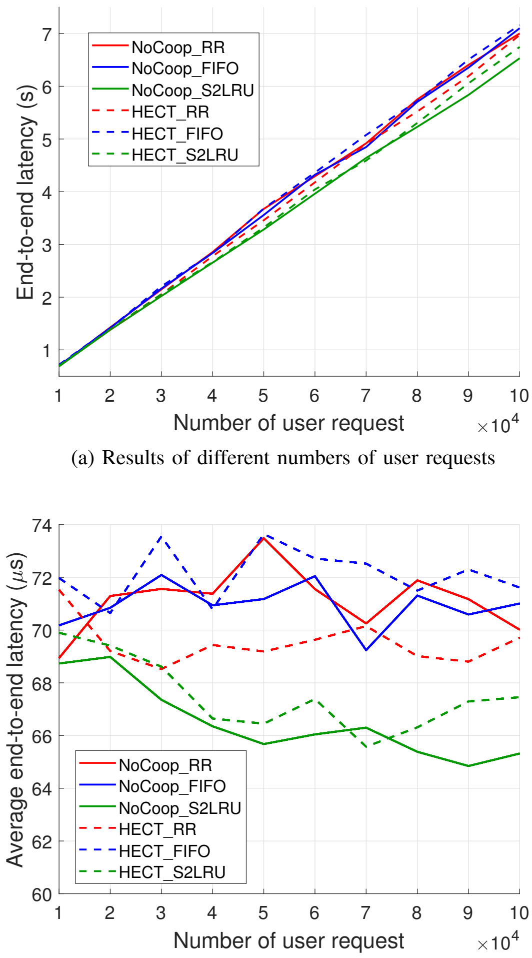
After a user u sending out a request req, we may prepare for the time-saving and energy efficient caching service from the perspective of the overall network provider. The most basic priority principle is the relative physical distance from users. First, check if the wanted file is already in local storage. Once it can be satisfied, we regard this case as no transmission latency. Second, req will be forwarded to nearby users within the scope of D2D communication. Third, when no cache can be found in any user, we may seek help from SBS Tier. Line 14 in Algorithm 2 includes the procedure of routing among SBSs under the signal range of the same MBS. In this way we can limit the routing times and avoid meaningless multi-hop forwarding. Fourth, once SBS caching is also failed, we need to upload the request to MBS Tier. Finally, if user request can not be satisfied by cache, we have to seek help from cloud server. Together with cache replacement policy shown in Algorithm 1, we may leave a copy at the each node of the route while downloading the requested file. Time complexity of Algorithm 2 is $O(n_{d2d} + n_{sbs})$.

V. Simulation and Analysis

In this section, we carry out experimental simulations to compare and analyze the performance of our proposed cache replacement policy and hybrid edge caching scheme with existing methods.

As shown in TABLE II, there are 1000 users, 100 small cells and 4 macro cells in a rectangular open area. In CL there are 200 files. The cache capacities of users, SBSs and MBSs are 2, 10, 50 MB. Sizes of the files in CL are uniformly distributed random numbers in the interval (0,1) MB. Signal range radius of D2D communication between users is 20 m, and radius of small cell is 100 m. According to the IMT-2020 5G specifications [40] which gives the peak data rate of 20 Gbps and user experienced data rate of 1Gbps in enhanced Mobile Broadband (eMBB) scenario, we set the 5G data rate in the simulation as 10Gbps. User power in D2D communication is 100 mW, which equals to 20 dBm. For SBS and MBS, 2 W and 100 W repectively equals to power level of 33 dBm and 50 dBm.

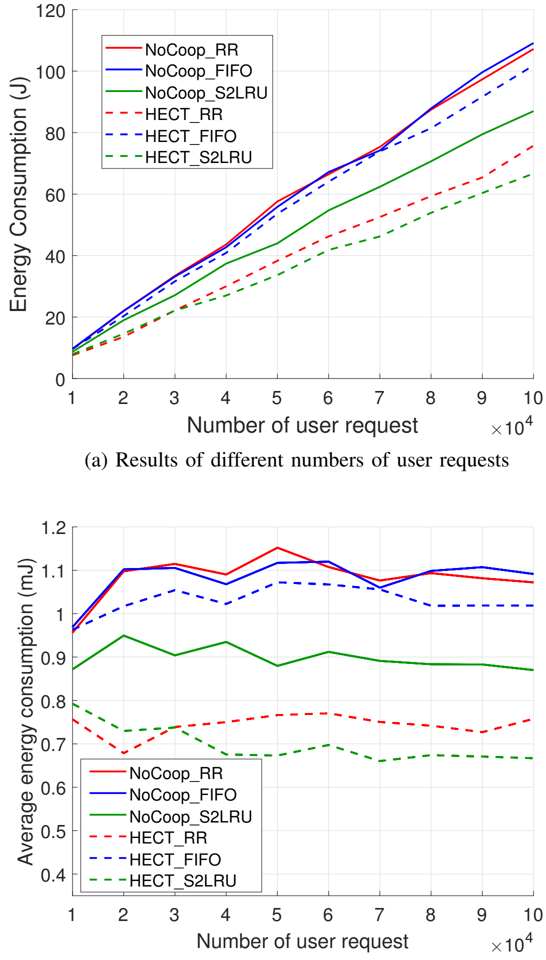
As comparison, we choose FIFO, Random Replacement (RR) and a NoHybd cache scheme. FIFO plays the role



(a) Results of different numbers of user requests

(b) Average results of single request

Fig. 5. Simulation results of end-to-end latency



(a) Results of different numbers of user requests

(b) Average results of single request

Fig. 6. Simulation results of energy consumption

of a low-overhead replacement policy, which can be easily achieved by setting FIFO queues. Facebook uses FIFO as its replacement policy [41]. RR is considered as a simple one which discards randomly and requires no access history. NoHybd stands for the traditional edge caching scheme. Regardless of the combination of the basic methods, NoHybd only caches during the vertical forwarding procedure. That is, no packet delivery happens among devices with the same network location including users and SBSs. We hope to use some extreme settings to evaluate the performance of cache replacement policies and cache schemes through multiple metrics, and provide valuable references for the realization of Tactile Internet in the near future.

A. Latency and Energy Consumption

Fig. 5 gives the simulation results of end-to-end latency. In Fig. 5a, we add together end-to-end latency of the ten groups in different user request numbers. First from the overall polyline trends of matches between replacement policy and cache scheme, the gaps are not large. Especially when number of requests is small, we can hardly judge the pros and cons of different methods through current experimental conditions. And as the number of user requests increases, there are gradually subtle changes. Caching schemes in S2LRU policy have certain advantages in numerical values. Actually our designed hybrid scheme is at a slight disadvantage in terms of latency which may be explained by the extra attempts looking for cached copies. That is, there exists no consumption on D2D communication and routing among SBSs under the same MBS.

Some further evaluation can be made from the Fig. 5b. We average the results of Fig. 5a by single request. As shown in this subfigure, polylines of NoHybd_RR, NoHybd_FIFO and HECT_FIFO are in a state of fluctuation as well as high value. HECT_RR is relatively stable, and lower in the overall value. Both schemes in S2LRU first decline and then stabilize as the number of user requests increases. As a result, we may judge that S2LRU can achieve better performance in reducing latency than regular replacement policies, especially when facing large request number.

Fig. 6 shows the results of energy consumption. Compared with Fig. 5, energy cost on processing requests from 1000 users show different patterns. HECT with three replacement policies saves more energy than NoHybd, respectively. HECT_FIFO still performs poorly, which can be explained that for fast-accumulated copies, always discarding the ones

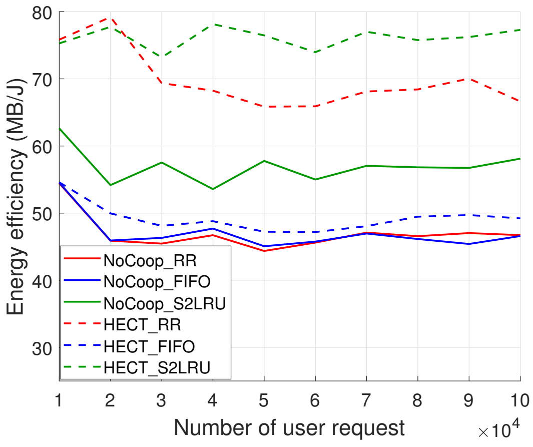


Fig. 7. Simulation results of energy efficiency

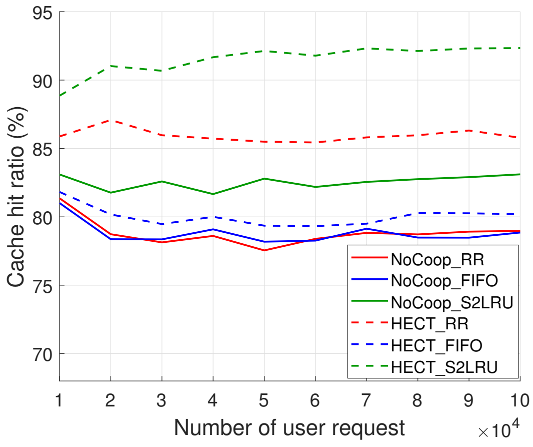


Fig. 8. Simulation results of cache hit ratio

first coming in is not only unable to take care of the popular files, but may also be counterproductive. In particular, as edge devices do not have much capacity for caching, files prioritized in Zipf distribution may be thrown away before giving preferential treatment. HECT_RR shows unexpected fair performance. Random eviction here gives the same level of TTLs to files with different popularities from the aspect of cache replacement.

B. Energy Efficiency and Cache Hit Ratio

According to Equation (12) and (14), Fig. 7 and 8 show the results of energy efficiency (EE) and cache hit ratio.

In pursuit of doing the most with the least amount of energy, we need to make reasonable scheduling of limited resources. For the problem of edge caching, we leave the copies on the file delivery route back to users and make efforts on finding needed copies for each user as near as possible. In Fig. 7, the maximal gap exists between HECT_S2LRU and NoHybd_RR/NoHybd_FIFO. In numerical value, for each Joule we may expect delivering 30 more MB with HECT_S2LRU than NoHybd_RR/NoHybd_FIFO. Since EE is in relation to latency and energy consumption, we can obtain some similar analysis results. FIFO shows small difference in two edge caching schemes.

We add cache hit ratio to show the percentage that requests are satisfied by cache. The order of combinations of two cache schemes and three replacement policies is about the same with EE. The exception happens when request number is no more than 30,000, HECT_RR and HECT_S2LRU are numerically close to each other in EE. This may be explained by the greater uncertainty of RR with small number of repeated trials. As a result, for HECT_S2LRU under the current experimental settings, more than 90% of user requests can be satisfied by cache.

C. Energy Efficiency of Each File in Zipf Distribution

After the discussion of latency, energy consumption, efficiency and cache hit ratio in the face of different numbers of user requests, lastly we focus on the EE of each file in CL that obeys Zipf distribution.

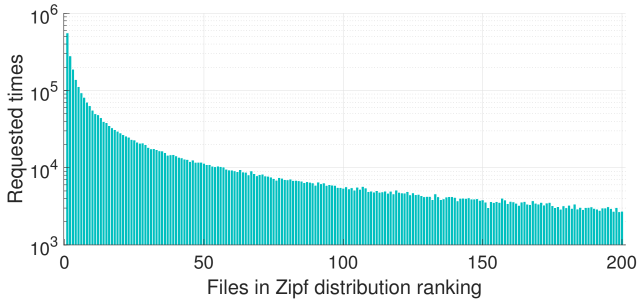
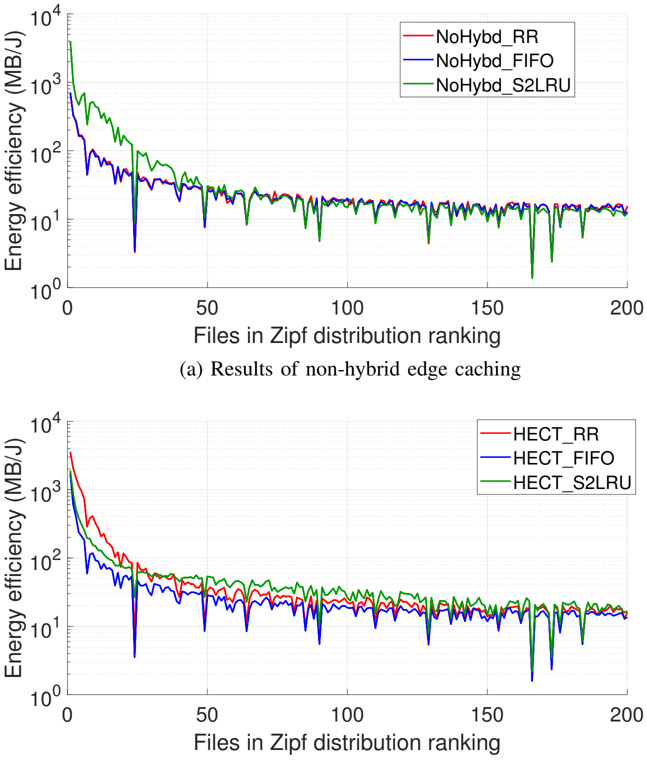


Fig. 9. Requested times of files in Zipf distribution ranking

Fig. 9 gives the total number of times each file requested in simulation. We set x in Equation (1) to 1 which is the classic version. And to display the values more intuitively, we change the y-axis to an exponential coordinate with a base of 10. Files in the top of Zipf ranking account for most of the total number (550,000) of requests. And files in bottom of ranking are close to each other.

Then we calculate the energy efficiency results different matches of replacement policies and edge cache schemes. Fig. 10a gives the EE of all 200 files in Zipf ranking. Two polylines of RR and FIFO are almost completely coincident. S2LRU in green achieve higher EE for the first 50 files in the front which may explain why the green solid line representing NoHybd_S2LRU can be ahead of the other three in Fig. 7. Lastly Fig. 10b displays the results of HECT. The order of three replacement policies is S2LRU, RR and FIFO, however, the first 30 files in the front obtains higher EE in RR. For each cached file, we win the opportunity to extend the TTL in segmented LRU. And this may partly lead to some uniformization of the treatment of files in the Zipf distribution. That is, the copies of a file in the top 30 may live no much longer than a top 31~60 (here we discuss the TTL of any single file copy). As a result, together with the hybrid edge caching scheme, S2LRU is able to give more attention to the files not in the front.

In summary, in this section, we prove through simulation



(a) Results of non-hybrid edge caching

(b) Results of hybrid edge caching

Fig. 10. Average energy efficiency of files in Zipf distribution ranking

results that the proposed replacement policy and cache scheme can improve the network performance in latency, energy consumption and energy efficiency from multiple aspects.

VI. Conclusion

In this paper, we focus on reducing latency and improving the energy efficiency of proactive in-network caching for the upcoming Tactile Internet in 5G. We analyze and summarize the existing caching methods and design a 3-tier network model. We first propose a replacement policy to set rules for cache eviction considering the popularity of cached files which obeys Zipf distribution. Then we put forward a hybrid edge caching scheme to integrate four existing ones taking effect in different parts of the network. The simulation results show that our methods can reduce latency and achieve better performance in overall energy efficiency.

Acknowledgment

This work is partially supported by JSPS KAKENHI Grant Number JP16K00117, and KDDI Foundation. Mianxiong Dong is the corresponding author.

References

[1] The Economist, "The next generation of wireless technology is ready for take-off," [Online]. Available: www.economist.com/business/2018/02/08/the-next-generation-of-wireless-technology-is-ready-for-take-off/, accessed July, 2018.

[2] M. Simsek, A. Aijaz, M. Dohler, J. Sachs and G. Fettweis, "5G-Enabled Tactile Internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460-473, March, 2016.

[3] CableFree, "LTE Network Latency compared with 2G, 3G & WiFi," [Online]. Available: www.cablefree.net/wirelesstechnology/4glte/lte-network-latency/, accessed July, 2018.

[4] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat and H. Dai, "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," *IEEE Communications Surveys Tutorials*, 2018. doi: 10.1109/COMST.2018.2841349.

[5] European Telecommunications Standards Institute, "Mobile Edge Computing - Introductory Technical White Paper," *Technical White Paper*, 2014.

[6] 5G Infrastructure PPP Association, "5G Vision-The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services." *Technical White Paper*, February, 2015.

[7] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta and P. Popovski, "Five Disruptive Technology Directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74-80, February, 2014.

[8] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour and G. Wunder, "5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201-1221, June, 2017.

[9] M. Chen, Y. Hao, M. Qiu, J. Song, D. Wu and I. Humar, "Mobility-Aware Caching and Computation Offloading in 5G Ultra-Dense Cellular Networks," *Sensors*, vol. 16, no. 7, 2016.

[10] M. Gregori, J. Gomez-Vilardebo, J. Matamoros and D. Gunduz, "Wireless Content Caching for Small Cell and D2D Networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1222-1234, May, 2016.

[11] M. Maier, M. Chowdhury, B. P. Rimal and D. P. Van, "The Tactile Internet: Vision, Recent Progress, and Open Challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 138-145, May, 2016.

[12] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos and M. Frodigh, "Realizing the Tactile Internet: Haptic Communications over Next Generation 5G Cellular Networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 82-89, April, 2017.

[13] H. Zhang, N. Liu, X. Chu, K. Long, A. H. Aghvami and V. C. M. Leung, "Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenge," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138-145, 2017.

[14] K. Antonakoglou, X. Xu, E. Steinbach, T. Mahmoodi and M. Dohler, "Towards Haptic Communications over the 5G Tactile Internet," *IEEE Communications Surveys Tutorials*, 2018. doi: 10.1109/COMST.2018.2851452.

[15] E. Ahmed and M. H. Rehmani, "Mobile Edge Computing: Opportunities, Solutions, and Challenges," *Future Generation Computer Systems*, vol. 70, pp. 59-63, 2017.

[16] M. Tao, K. Ota and M. Dong, "Foud: Integrating Fog and Cloud for 5G-Enabled V2G Networks," *IEEE Network*, vol. 31, no. 2, pp. 8-13, March, 2017.

[17] L. Li, K. Ota and M. Dong, "Deep Learning for Smart Industry: Efficient Manufacture Inspection System With Fog Computing," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4665-4673, October, 2018.

[18] L. Li, K. Ota and M. Dong, "DeepNFV: A Lightweight Framework for Intelligent Edge Network Functions Virtualization," *IEEE Network*, vol. 33, no. 1, pp. 136-141, January, 2019.

[19] Y. Mao, C. You, J. Zhang, K. Huang and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2322-2358, 2017.

[20] R. Ford, A. Sridharan, R. Margolies, R. Jana and S. Rangan, "Provisioning Low Latency, Resilient Mobile Edge Clouds for 5G," *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 169-174, May, 2017.

[21] H. Li, K. Ota and M. Dong, "Learning IoT in Edge: Deep Learning for the Internet of Things with Edge Computing," *IEEE Network*, vol. 32, no. 1, pp. 96-101, January, 2018.

[22] 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 Communications Magazine*, vol. 52, no. 2, pp. 131-139, February, 2014.

[23] E. Bastug, M. Bennis and M. Debbah, "Living on the Edge: The Role of Proactive Caching in 5G Wireless Networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82-89, August, 2014.

[24] H. Li, K. Ota and M. Dong, "ECCN: Orchestration of Edge-Centric Computing and Content-Centric Networking in the 5G Radio Access Network," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 88-93, June, 2018.

[25] M. Erol-Kantarci and S. Sukhmani, "Caching and Computing at the Edge for Mobile Augmented Reality and Virtual Reality (AR/VR) in 5G," *Ad Hoc Networks*, pp. 169-177, 2018.

[26] D. Li, M. Dong, Y. Yuan, J. Chen and K. Ota, "SEER-MCache: A Prefetchable Memory Object Caching System for IoT Real-Time Data Processing," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3648-3660, October, 2018.

[27] T. X. Tran, A. Hajisami, P. Pandey and D. Pompili, "Collaborative Mobile Edge Computing in 5G Networks: New Paradigms, Scenarios, and Challenges," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 54-61, April, 2017.

[28] J. Kwak, Y. Kim, L. B. Le and S. Chong, "Hybrid content caching for low end-to-end latency in cloud-based wireless networks," *2017 IEEE International Conference on Communications (ICC)*, pp. 1-6, May, 2017.

[29] J. Kwak, Y. Kim, L. B. Le and S. Chong, "Hybrid Content Caching in 5G Wireless Networks: Cloud Versus Edge Caching," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3030-3045, May, 2018.

[30] N. Deng and M. Haenggi, "The Benefits of Hybrid Caching in Gauss-Poisson D2D Networks," *IEEE Journal on Selected Areas in Communications*, pp. 1-1, 2018.

[31] D. M. W. Powers, "Applications and Explanations of Zipf's Law," *Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning*, pp. 151-160, January, 1998.

[32] R. Rousseau, "George Kingsley Zipf: life, ideas, his law and informetrics," *Glottometrics*, vol. 3, no. 1, pp. 11-18, 2002.

[33] L. A. Adamic and B. A. Huberman, "Zipf's law and the Internet," *Glottometrics*, vol. 3, no. 1, pp. 143-150, 2002.

[34] J. G. Andrews, F. Baccelli and R. K. Ganti, "A Tractable Approach to Coverage and Rate in Cellular Networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122-3134, November, 2011.

[35] X. Ge, B. Yang, J. Ye, G. Mao, C. Wang and T. Han, "Spatial Spectrum and Energy Efficiency of Random Cellular Networks," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 1019-1030, March, 2015.

[36] A. Zappone and E. Jorswieck, "Energy Efficiency in Wireless Networks via Fractional Programming Theory," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 3-4, pp. 185-396, 2015.

[37] S. Buzzi, C. I, T. E. Klein, H. V. Poor, C. Yang and A. Zappone, "A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 697-709, April, 2016.

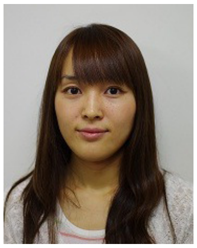
[38] F. Gabry, V. Bioglio and I. Land, "On Energy-Efficient Edge Caching in Heterogeneous Networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3288-3298, December, 2016.

[39] A. Zappone, E. Björnson, L. Sanguinetti and E. Jorswieck, "Globally Optimal Energy-Efficient Power Control and Receiver Design in Wireless Networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2844-2859, June, 2017.

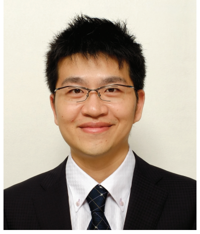
[40] Series M, IMT Vision–Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond. *Recommendation ITU-R M.2083-0*, September, 2015.

[41] Q. Huang, K. Birman, R van Renesse, W. Lloyd, S. Kumar and H. C. Li, "An Analysis of Facebook Photo Caching," *Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles (SOSP '13)*, pp. 167-181, 2013.

Jianwen Xu received the B.Eng degree in Electronic and Information Engineering from Dalian University of Technology (DLUT), China, in 2014, and M.Eng degree in Information and Communication Engineering from Shanghai Jiao Tong University (SJTU), China, in 2017. He is currently pursuing the Ph.D. degree in Electrical Engineering at Muroran Institute of Technology, Japan. He is selected as a Non-Japanese Researcher (a total of 3 recipients all over Japan) by NEC C&C Foundation for 2019 Fiscal Year. His main fields of research interest include distributed system, Internet of things.



Kaoru Ota was born in Aizu-Wakamatsu, Japan. She received M.S. degree in Computer Science from Oklahoma State University, USA in 2008, B.S. and Ph.D. degrees in Computer Science and Engineering from The University of Aizu, Japan in 2006, 2012, respectively. She is currently an Assistant Professor with Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. From March 2010 to March 2011, she was a visiting scholar at University of Waterloo, Canada. Also she was a Japan Society of the Promotion of Science (JSPS) research fellow with Kato-Nishiyama Lab at Graduate School of Information Sciences at Tohoku University, Japan from April 2012 to April 2013. Her research interests include Wireless Networks, Cloud Computing, and Cyber-physical Systems. Dr. Ota has received best paper awards from ICA3PP 2014, GPC 2015, IEEE DASC 2015, IEEE VTC 2016-Fall, FCST 2017, 2017 IET Communications Premium Award and IEEE ComSoc CSIM Best Conference Paper Award 2018. She is an editor of IEEE Transactions on Vehicular Technology (TVT), IEEE Communications Letters, Peer-to-Peer Networking and Applications (Springer), Ad Hoc & Sensor Wireless Networks, International Journal of Embedded Systems (Inderscience) and Smart Technologies for Emergency Response & Disaster Management (IGI Global), as well as a guest editor of ACM Transactions on Multimedia Computing, Communications and Applications (leading), IEEE Internet of Things Journal, IEEE Communications Magazine, IEEE Network, IEEE Wireless Communications, IEEE Access, IEICE Transactions on Information and Systems, and Ad Hoc & Sensor Wireless Networks (Old City Publishing). She is the recipient of IEEE TCSC Early Career Award 2017, and The 13th IEEE ComSoc Asia-Pacific Young Researcher Award 2018.



Mianxiong Dong received B.S., M.S. and Ph.D. in Computer Science and Engineering from The University of Aizu, Japan. He is currently an Associate Professor in the Department of Information and Electronic Engineering at the Muroran Institute of Technology, Japan. He was a JSPS Research Fellow with School of Computer Science and Engineering, The University of Aizu, Japan and was a visiting scholar with BBCR group at University of Waterloo, Canada supported by JSPS Excellent Young Researcher Overseas Visit Program from April 2010 to August 2011. Dr. Dong was selected as a Foreigner Research Fellow (a total of 3 recipients all over Japan) by NEC C&C Foundation in 2011. His research interests include Wireless Networks, Cloud Computing, and Cyber-physical Systems. He has received best paper awards from IEEE HPCC 2008, IEEE ICESS 2008, ICA3PP 2014, GPC 2015, IEEE DASC 2015, IEEE VTC 2016-Fall, FCST 2017, 2017 IET Communications Premium Award and IEEE ComSoc CSIM Best Conference Paper Award 2018. Dr. Dong serves as an Editor for IEEE Transactions on Green Communications and Networking (TGCN), IEEE Communications Surveys and Tutorials, IEEE Network, IEEE Wireless Communications Letters, IEEE Cloud Computing, IEEE Access, as well as a leading guest editor for ACM Transactions on Multimedia Computing, Communications and Applications (TOMM), IEEE Transactions on Emerging Topics in Computing (TETC), IEEE Transactions on Computational Social Systems (TCSS). He has been serving as the Vice Chair of IEEE Communications Society Asia/Pacific Region Information Services Committee and Meetings and Conference Committee, Leading Symposium Chair of IEEE ICC 2019, Student Travel Grants Chair of IEEE GLOBECOM 2019, and Symposium Chair of IEEE GLOBECOM 2016, 2017. He is the recipient of IEEE TCSC Early Career Award 2016, IEEE SCSTC Outstanding Young Researcher Award 2017, The 12th IEEE ComSoc Asia-Pacific Young Researcher Award 2017, Funai Research Award 2018 and NISTEP Researcher 2018 (one of only 11 people in Japan) in recognition of significant contributions in science and technology. He is currently the Member of Board of Governors and Chair of Student Fellowship Committee of IEEE Vehicular Technology Society, and Treasurer of IEEE ComSoc Japan Joint Sections Chapter.