Proactive Cache System Using Cellular-Radio Information on MEC

Miho Tanaka*, Kenichi Nagasue*, Jun Ogawa*, Akira Yokomichi†, Yoshihiro Ando†,
Takayuki Ojima‡, Toshiyuki Shizuoka‡, Koichi Adachi‡, Takeo Fujii‡
* FUJITSU LABORATORIES LTD., Japan
{miho.tanaka, nagasue.kenichi, ogawa.jun}@fujitsu.com
† FUJITSU LIMITED, Japan
{yokomichi.akira, ando.yoshihiro}@fujitsu.com
‡ Advanced Wireless & Communication Research Center(AWCC), Japan
{ojima, shizuoka, adachi, fujii}@awcc.uec.ac.jp

***Abstract*—To achieve a quick response to mobile terminals and utilize the large volumes of data in mobile network, it is proposed to use resources of edges of mobile network as known as Multi-Access Edge Cloud (MEC) and integrate MEC with clouds. A proactive cache technology, which allocates cache data to MEC servers accessed by mobile terminals in advance, is one of technologies to provide the data in a low latency for the mobile terminals. The technology, however, less considers both of the cache placement in time for handover (HO) of mobile terminals and the increase in the traffic between MEC servers for cache distribution. The increase in the traffic is caused by duplicated routes of cache transfer from a location of content data to cache locations and by cache placements on more MEC servers to improve cache hit rate according to a mobile terminal transition. In this paper, we propose a novel proactive cache system which transfers cache data in time for HO of mobile terminals and reduces the traffic between MEC servers, using a route de-duplicating technique for cache transfer and selective data delivery utilizing the locations of mobile terminals predicted from cellular-radio information. We also propose the system architecture which enables services to utilize cellular-radio information by APIs, and evaluate the system by a simulation of the HO prediction and the estimation of the traffic between MEC servers for transferring caches. Numerical results indicate 49.1% reduction of the total traffic between MEC servers and reveal characteristics and problems of the cache system using cellular-radio information.**

***Index Terms*—Multi-Access Edge Computing, edge computing, mobility, cache system, handover prediction, cellular radio**

I. Introduction

Along with the increase in the number of mobile terminals, the volume of data generated or consumed by the mobile terminals in the mobile network becomes larger. Ericsson shows that global mobile data was as much as about 20 exabytes (EB) per month in 2018 and 85% of the data was generated by smartphones [1]. An amount of the data will increase and reach close to 107 EB per month by 2023.

With the increase in mobile network data, new services using these data can be provided to mobile terminals. For example, AR/VR services will be provided from the edge of the mobile network [2]. In addition to this, a video sharing service like a real-time street view service will also be provided using the video generated by mobile terminals. These services provide the large volumes of data and a quick response for better user experience utilizing the short distance between a mobile terminal and the edge.

To realize these services, the integration of cloud and edges of the mobile network is proposed against cloud-intensive systems. For a quick response, the European Telecommunications Standards Institute (ETSI) promotes the standardization of Multi-Access Edge Computing (MEC) technology, which stores or processes data generated by mobile terminals on the servers located near base stations (MEC servers) [3]. Since computing or processing large volumes of data requires more resources on the cloud than MEC, Fujitsu's Multi-Access Edge Cloud, based on the ETSI's MEC, proposes to use the cloud and MEC properly according to the requirements of services [4]. There are also some services and concepts that propose the integration by Cisco's Fog Computing [5], Ericsson's Distributed Cloud [6] and Nokia's MEC platform [7].

To deal with large volumes of data on MEC servers and to achieve mobility responsiveness, a proactive cache [8] is one of the major approaches. The existing proactive cache approach selects the MEC servers which will be connected from mobile terminals in the next and delivers caches to the selected ones before accessed from mobile terminals. It benefits when a mobile terminal moves to a new MEC server and gets the content data from the local MEC servers immediately after handover (HO).

However, with current proactive cache approach the traffic between MEC severs for the cache distribution will increase in the following cases. First, as the number of located cache data increases according to the number of mobile terminals, the transfer routes duplicate more. The cache approach considers the network topology between the cache location and the location of mobile terminal for low latency, but it less considers between the content data location and cache locations for the traffic. Second, the more number of the cache data are located than the actual number of the cache data accessed by mobile terminals in order to improve a cache hit rate. These cases increase the traffic while network resource between MEC servers are more limited than the cloud.

In the context of content cache placements, there are many

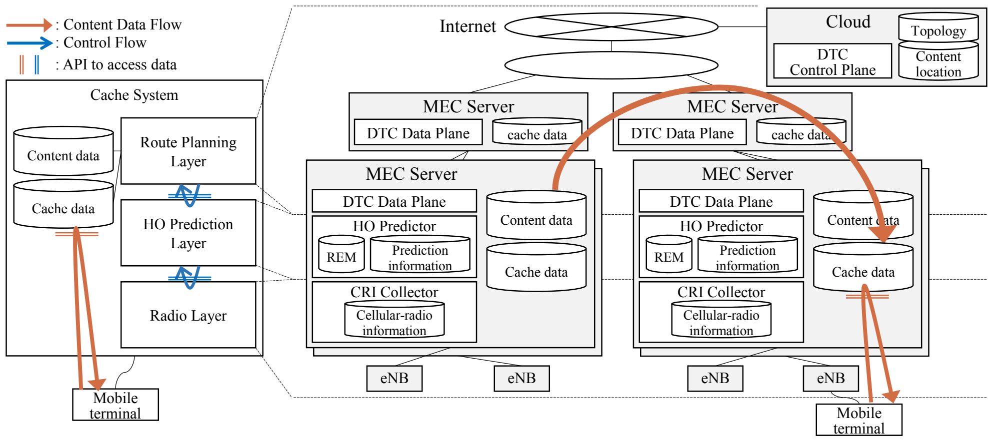


Fig. 1: Proactive cache system architecture using cellular-radio information

previous studies but they less consider the increase in the traffic for the cache distribution. [9] considers network topology to distribute caches for reduction of network traffic, which focuses on the efficiency between cache locations and mobile terminals. [8] and [10] also consider topology for a lower latency to mobile terminals and a limit of storage, not for reducing the duplicated cache transfers. [11] uses pre-prepared data popularity for improving a cache hit rates and introduces the concept of divisions of sequential data. [12] also takes a mobility-based cache policy to decide the cache location, but supposes to use all history of mobile terminal's transition.

To address the increase in the traffic, we propose a novel proactive cache system which reduces the amount of the traffic of cache delivery between MEC servers, using a route deduping technique for data transfer and selective data delivery utilizing the locations of mobile terminals predicted from cellular-radio information. The prediction using the information not only enables a reduction in the number of MEC servers to be located cache data, but also provides the cache system without GPS of mobile terminals when they use the data. Additionally, the proposed system eliminates the duplicated routes of transfers by calculating the route and the order to transfer caches onto MEC servers. The proposed system also enables to keep to transfer cache data in time for HO of mobile terminals using the HO prediction while reducing the traffic.

The rest of this paper is organized as follows: in section II we show an architecture of the proposed cache system. Section III describes the algorithm that predicts the HO of the mobile terminal and Section IV shows the algorithm to calculate the route to each delivery cache. Section V and Section VI present the evaluation and discussion on the proposed system. In Section VII we conclude the paper.

II. Proactive Cache System Architecture

A. Architecture

Fig. 1 shows the architecture of the proposed cache system, which is composed of three layers, Radio layer, HOP layer, and RP layer.

- Radio layer: This layer provides functions of eNB and cellular-radio information (CRI) Collector which collects pathloss. In order to collect the information, the functions are used a software-based 3GPP cellular network based on OpenAirInterface (OAI) [13]. The collected information can be accessed by APIs.
- HO Prediction layer (HOP layer): In this layer, HO Predictor predicts MEC servers accessed by a mobile terminal after the next HO and the HO time using the cellular-radio information. The detail of HO Predictor is described in section III.
- Route-Planning layer (RP layer): In this layer, Data Transfer Controller (DTC) duplicates the transfer route of cache data to reduce cache delivery traffic. The detail of DTC is described in section IV.

The proposed system consists of eNB, MEC server and cloud. Content data or cache data is stored on each MEC server and mobile terminals can access the data on the nearest MEC server through eNB. The functions in the proposed system run at constant time intervals in accordance with the changes of MEC servers accessed by mobile terminals with time. The structure enables mobile terminals to access cache data on the nearest MEC server immediately after HO, since DTC transfers cache data to the predicted MEC servers in advance.

B. Use-case

As an effective use-case applied to the proposed system, we introduce a service: a real-time street view service in Fig. 2. This service uses videos which are took by mobile terminals on street and stored on MEC servers. The other mobile terminals require the video to watch the condition on their direction. Videos of the street easily tell them more detailed information such as their speed, brake light and weather than pictures.

The proposed system transfers the cache data of the video required by mobile terminals on the MEC server which will be accessed by them in Fig. 2. The placement before their

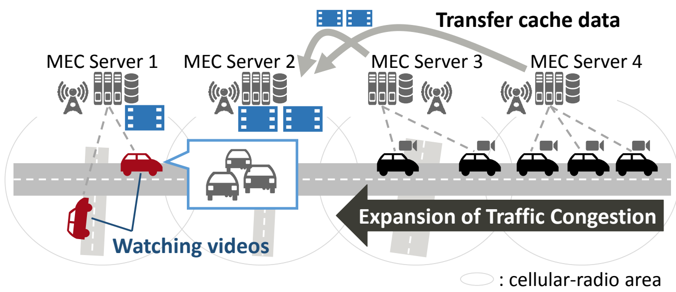


Fig. 2: Use-case: real-time street view

HO enables the mobile terminal to watch the video without breaking off after transition between MEC servers.

III. HandOver (HO) Prediction Based on Distance Estimation Algorithm Using REM

The proposed handover (HO) prediction algorithm utilizes information extracted from a radio environment map (REM). The REM is constructed based on the information such as the mobile terminal's position and received signal strength (RSS) in advance [14], [15]. By comparing the received signal power measured by a mobile terminal with what is stored in REM, it is possible to estimate the position of the mobile terminal. The HO timing prediction is performed based on the mobile terminal's location history.

A. REM Construction

In this section, the construction of REMs is briefly explained. Fig. 3 shows the concept of REMs. By creating an REM, it is possible to understand the radio environment at a specific position. For example, an REM of the average received power observed by a mobile terminal can provide information about the average power received from an evolved NodeB (eNB) at a specific position. During the construction of an REM, a mobile terminal equipped with GPS is used. In this research, it is assumed that REM is placed at an MEC server. Thus, the radio environment information such as the transmitter position, the received signal strength (RSS) at eNB, and the center frequency are collected at eNB and stored in the database. The observation area is divided into a two-dimensional mesh, and the measured information at each mesh is statistically processed in order to generate an REM. An average received power observed by a mobile terminal at each mesh is stored in the REM. Thus, during the operation phase of REM, no information is collected from mobile terminal except for information such as RSS.

B. Distance Estimation Algorithm Using REM

In order to predict the HO timing, estimation of the distance between a terminal and an eNB is performed. For this estimation, GPS is not used. In this research, the whole simulation area is divided into L-by-L [m^2] meshes. Comparing the stored value at each mesh of REM and the observation value, the

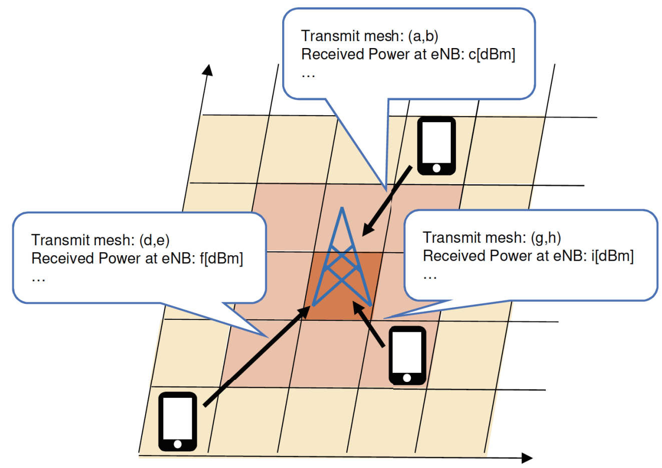


Fig. 3: Example of REM construction

candidate meshes where the terminal is currently residing in are estimated.

Without loss of generality, let us consider the kth distance estimation. Let a two dimensional index of a mesh be denoted by (i, j) where i and j are the indices of the x-axis and y-axis, respectively. The statistical value of RSS at mesh (i, j) is expressed as $P(i, j)$ [dBm]. The distance estimation is carried out every τ [sec], thus the time at which the kth distance estimation is performed is expressed as:

$$t_k = t_{k-1} + \tau. \tag{1}$$

Let us denote the number of RSS values measured between $[t_{k-1}, t_k)$ with L_k. The kth distance estimation is performed as follows:

1) The average RSS value P_k [dBm] is obtained by averaging L_k measurements between $[t_{k-1}, t_k)$.
2) Create a set of meshes $\mathcal{A}_k$ as:
$$\mathcal{A}_k = \{(i, j) \mid P_k - \Delta \leq P(i, j) < P_k + \Delta\}, \tag{2}$$
where Δ [dB] is a margin.
3) The following expanded set $\mathcal{B}_k^{(1)}$ is created by selecting the mesh $(i, j) \in \mathcal{A}_k$ and its surrounding eight meshes as:
$$\mathcal{B}_k^{(1)} = \{(i \pm 1, j \pm 1) \mid (i, j) \in \mathcal{A}_k\}. \tag{3}$$
4) Set $n = 1$ and repeat the following steps 5 and 6 until n reaches N.
5) Obtain the product set $C_k^{(n+1)} = \mathcal{B}_k^{(n)} \cap \mathcal{A}_{k-n}$.
6) Create the expanded set $\mathcal{B}_k^{(n+1)}$ as:
$$\mathcal{B}_k^{(n+1)} = \left\{(i \pm 1, j \pm 1) \mid (i, j) \in C_k^{(n+1)}\right\}, \tag{4}$$
and increment n by 1.
7) The distances between the eNB and the meshes in the set C_k^N are calculated. Averaging the calculated distances gives the estimated average distance d_k.
8) In order to smooth the estimated distance, a forgetting factor $\gamma \in (0, 1)$ is introduced as:
$$d'_k = (1 - \gamma)d_k + \gamma d'_{k-1}. \tag{5}$$

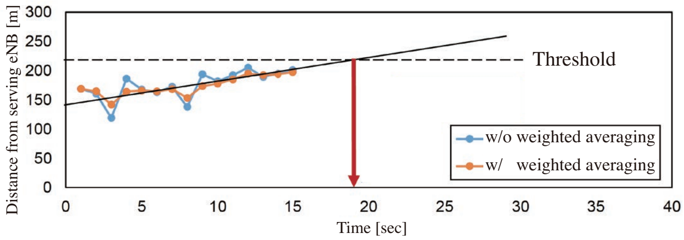


Fig. 4: HO timing prediction

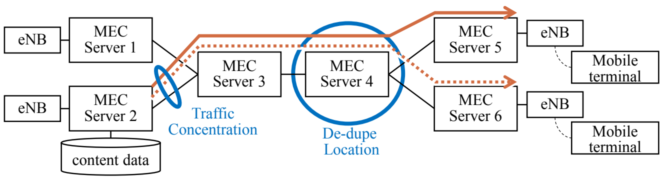


Fig. 5: Elimination of duplicated transfer between MEC servers

C. HO Timing Prediction Using Estimated Distance

HO happens when the terminal is approaching to the boundary of the eNB's coverage. It means that if the distance d'_k is close to the coverage of the eNB D, HO happens with high probability. By approximating d'_k as a straight line by linear regression, it is possible to estimate the HO timing t_{HO}. The HO timing t_{HO} is given by the intersetion of the approximated line and threshold D as shown in Fig. 4.

D. Direction Prediction of HO

Based on the remaining meshes in Subsection III-B, the HO probability to each eNB is derived. The coverage area of eNB is divided into non-overlapping regions according to the number of neighboring eNBs, and each region is associated with a specific neighboring eNB. Note that the division into non-overlapping regions should be determined based on the number of neighboring eNBs and the cell shape. Suppose there are B neighboring eNBs and the number of remaining meshes in each region associated with the bth neighboring eNB is m_b. Then, the HO probability to the bth neighboring eNB is given by:

$$p_b = \frac{m_b}{\sum_{b'=1}^{B} m_{b'}}. \quad (6)$$

IV. Route Planning Algorithm for Transferring Caches

In this section we describe an algorithm that reduces the traffic for cache distributions between MEC servers. The DTC is separated into two functions, Control Plane on cloud and Data Plane on MEC servers. The algorithm runs on the control plane of the DTC on cloud, and de-duplicates the routes to transfer content data to mobile terminals.

Fig. 5 shows an image to reduce the traffic between MEC servers. When some mobile terminals request one shared content data on MEC Server 2, the traffic to transfer the data concentrates at the MEC server storing the data. The traffic load increases in accordance with the number of the requesting mobile terminals. Our idea to reduce the traffic load is simple. DTC finds de-duplicating locations where the shortest routes from the source MEC server to each requiring mobile terminal branch off like MEC Server 4, because locating a cache on the de-duplicating location collects the requests together at the cache data and eliminates duplicated transfer. The MEC servers which will be accessed from mobile terminals are given from the HO Predictor as results of the HO prediction. The traffic between MEC servers is influenced by the precision of the HO Predictor and the number of MEC servers.

The following shows processes of DTC. DTC uses a network topology, locations of content data and predicted information of the requiring mobile terminals as inputs. First, DTC on cloud calculates the shortest routes from the source MEC server to each requiring mobile terminal. Second, DTC finds the de-duplicating location for each mobile terminal. The outputs of the calculations are cache locations of MEC servers which are predicted to be accessed directly by mobile terminals next and MEC servers which eliminate duplicated transfer, and the order to transfer the cache data. Then, DTC on cloud instructs the DTC Data Plane on MEC servers to transfer cache according to the result.

V. Numerical Evaluation

A. Target of Evaluation

The target of this evaluation is to measure the efficiency of traffic reduction based on cellular-radio information and network topology, and to reveal problems derived from the characteristics of cellular-radio information to cache system over MEC servers. First, we simulate the HO prediction in the simulation environment of one base station and show the characteristics of the prediction. Second, we estimate traffic between MEC servers by applying DTC and the simulated probabilities of the HO predictor to one use-case, and show the traffic reduction.

B. Simulation Parameters

In this section, we will provide the numerical results in order to show the performance of the proposed scheme. A carrier frequency of 900 [MHz] is considered. The Okumura-Hata model is adopted for distant dependent pathloss and the spatially correlated shadowing model with the standard deviation of 6 [dB] is considered for channel model. The speed of the mobile terminal is set to 1.11 [m/s] which corresponds to the walking speed. The mesh size of the REM is set to $L = 2$ [m]. It is assumed that HO occurs when the terminal arrives at a distance of 250 [m] from eNB. The HO prediction is carried out every $\tau = 1$ [sec]. In this simulation, the number of past information used for distance estimation algorithm is set to $K = 10$.

C. Traffic Estimation Parameters

The parameters for an estimation of traffic between MEC servers are decided by the use-case of the video sharing service

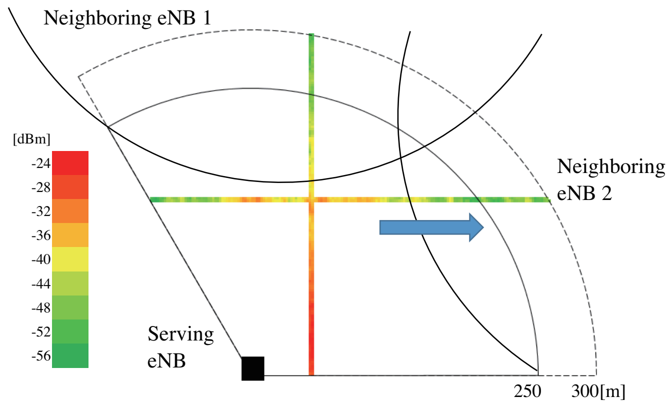


Fig. 6: Moving trajectory for performance evaluation.

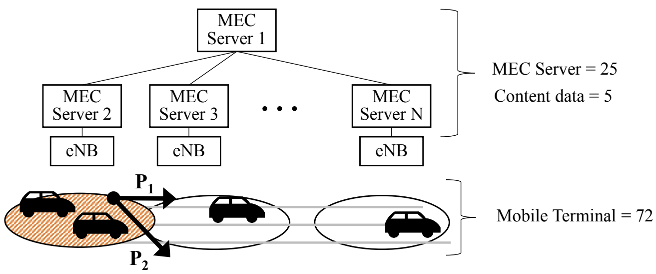


Fig. 7: Use-case of the estimation

which are shown in section I and Fig. 7. The DTC locates caches on 25 MEC servers for 72 mobile terminals. The number of mobile terminals is based on the traffic jam statistics in a unit distance and the ratio of the service users in whole mobile terminals. The cache data size per one mobile terminal is 100MB (two 4K video files for 10 seconds). We use the probabilities simulated in section III-D to express the behavior of mobile terminals under MEC servers: P_1: HO probability in the non-turning direction; P_2: HO probability in the turning direction.

In this estimation we compare the proposed cache system with a no-cache system, a naïve cache system which always places two cache data for both of directions, and an ideal cache system that doesn't have any miss predictions or any duplicated route transferring caches. The metrics are (a) a total amount of traffic of all connections and (b) an amount of traffic of a connection with the highest traffic load.

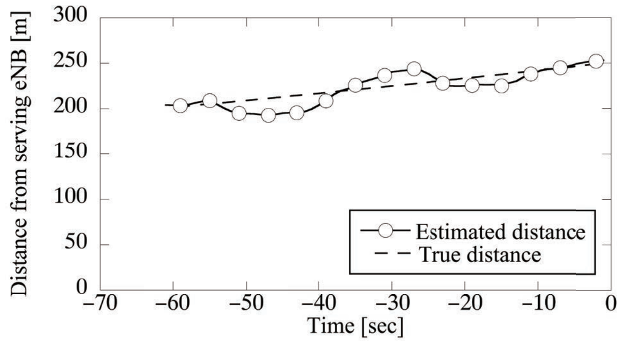
VI. Result and Discussion

The HO timing prediction under the simulation environment shown in Fig. 6 is considered. The eNB currently communicating with the mobile terminal is located at the center of the arc and there are two neighboring eNBs. The mobile terminal is assumed to be moving along the horizontal street. It is further assumed that the REM storing the average received power at each mesh on the streets has been perfectly created. Fig. 8a shows the estimated distance of the mobile terminal from the communicating eNB. As Fig. 8a shows the proposed algorithm can provide an estimated distance that is relatively close to the true one. Fig. 8b shows the absolute error value of the estimated HO timing. As it can be seen from the figure, as the mobile terminal approaches the coverage boundary, the estimation error becomes smaller. In this simulation, since we do not consider the effect of surrounding buildings, the site specific radio propagation is not emulated. Some degradation of distance estimation appears in Fig. 8a. The timing of HO shown in Fig. 8b is estimated with linear regression with past K estimated distances. Therefore, several 10 seconds error occurs. When the HO timing estimation starts, e.g., about -60 second, the prediction error is large due to the insufficient number of estimation history. As the number of estimation history increases, the estimation error becomes smaller. The reason why the estimation error becomes large between -20 to 0 second is the existence of the meshes that contains the similar RSS values. Due to those meshes, the proposed HO timing estimation algorithm output the wrong mobile terminal location. The HO probabilities for two neighboring eNBs are shown in Fig. 8c. Since the mobile terminal is moving along the horizontal street, it will be handed over to eNB 2. It is shown from the figure that the proposed algorithm can provide the correct HO probability.

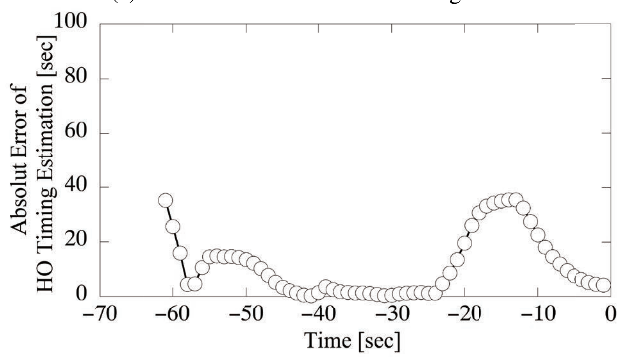
Fig. 9 shows the amount of traffic of each system. The proposed system shows 49.1% reduction of the traffic in (a) and the 55.3% reduction in (b) by the proposed cache system comparing the no-cache system. The proposed system, however, is 23.8% higher than ideal or naïve cache system in (b) because of the lower precision of the probabilities provided by the HO predictor and the cache transfer from a MEC server with content data to the accessed MEC server directly. The naïve system also transfers the more number of the cache data than the ideal system, but the amount of traffic has no influence of miss-prediction. It means that the precision has a large impact on traffic.

For the more traffic reduction, it needs the better precisions of the HO prediction and the more efficient cache placement by the DTC. We have two problems for the improvements as our future works as the followings:

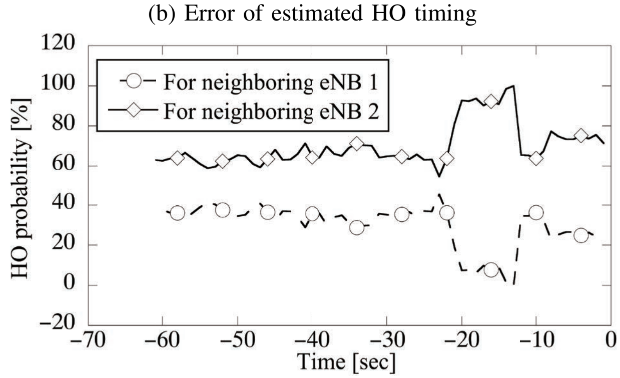
- (1) The DTC needs a cache replacement considering the difference of the radio characteristics. Since the characteristics of the average received power in each REM on MEC servers are different from the road or buildings under the base stations, the DTC cannot handle the predicted probability with uniform accuracy.
- (2) The DTC also needs to distribute caches for mobile terminals staying in a short time under one MEC server, combining the HO predictions from some MEC servers to know the movement tendency of the mobile terminal. The current system cannot transfer a cache to a MEC server for a mobile terminal with a short stay under that, because the HO prediction needs k distance estimations as a mobile terminal's history data.



(a) Estimated distance from serving eNB



(b) Error of estimated HO timing



(c) HO probabilities

Fig. 8: Performance of the proposed algorithm

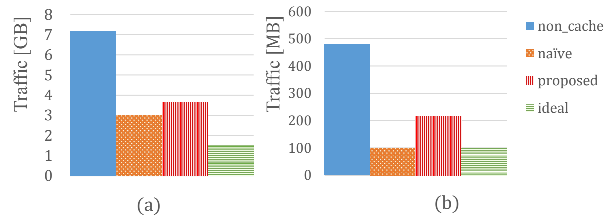


Fig. 9: Traffic for cache distribution between MEC servers. (a) Total amount of traffic of all connections. (b) Amount of traffic of a connection with the highest traffic load.

VII. Conclusion

In this paper we proposed a novel proactive cache system which uses cellular-radio information in order to reduce the traffic between MEC servers. The proposed system reduces the amount of the traffic by selecting MEC servers to be placed cache data using the HO prediction and by de-duplicating the routes to transfer caches. Using cellular-radio information for the HO prediction enables end-users to use the cache system without GPS and access data immediately after HO. The numerical result of the evaluations shows that the proposed system reduces the 49.1% of the total amount of the traffic between MEC servers compared to a system with no cache techniques and the system needs to be improved for more reduction according to radio characteristics and transition tendency of mobile terminals.

References

[1] Telefonaktiebolaget L. M. Ericsson, "Ericsson Mobility Report," June 2018, accessed Apr. 2019. [Online]. Available: https://www.ericsson.com/en/mobility-report/reports/q4-update-2018

[2] AT&T Inc., "AT&T Foundry Powers Up Edge Computing Test Zone in Silicon Valley to Drive Innovation in 5G Era," Feb 2018, accessed Apr. 15, 2019. [Online]. Available: https://about.att.com/story/att_foundry_powers_up_edge_computing_test_zone.html

[3] European Telecommunications Standards Institute, "Multi-access Edge Computing (MEC)," accessed Apr. 2019. [Online]. Available: https://www.etsi.org/technologies/multi-access-edge-computing

[4] A. Yoshihiro, O. Jun, U. Hitoshi, and T. Nobuhiro, "Fujitsu Multi-access Edge Cloud to Realize Diverse Digital Services in 5G Era," FUJITSU LIMITED, Tech. Rep., Nov 2018. [Online]. Available: https://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol69-6/paper07.pdf

[5] Cisco Systems, Inc., "Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are," 2015, accessed Apr. 15, 2019. [Online]. Available: https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf

[6] Telefonaktiebolaget L. M. Ericsson, "Going beyond edge computing with distributed cloud," accessed Apr. 15, 2019. [Online]. Available: https://www.ericsson.com/en/digital-services/trending/distributed-cloud

[7] Nokia Corporation, "Multi-access Edge Computing (MEC)," accessed Apr. 15, 2019. [Online]. Available: https://networks.nokia.com/solutions/multi-access-edge-computing

[8] V. A. Siris, X. Vasilakos, and G. C. Polyzos, "Efficient proactive caching for supporting seamless mobility," in *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, June 2014, pp. 1–6.

[9] T. Hou, G. Feng, S. Qin, and W. Jiang, "Proactive content caching by exploiting transfer learning for mobile edge computing," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–6.

[10] K. Zhang, S. Leng, Y. He, S. Maharjan, and Y. Zhang, "Cooperative content caching in 5g networks with mobile edge computing," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 80–87, JUNE 2018.

[11] A. Mahmood, C. Casetti, C. F. Chiasserini, P. Giaccone, and J. Harri, "Mobility-aware edge caching for connected cars," in *2016 12th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, Jan 2016, pp. 1–8.

[12] F. Zhang, , Y. Zhang, K. K. Ramakrishnan, S. Mukherjee, and R. Y. and, "Edgebuffer: Caching and prefetching content at the edge in the mobilityfirst future internet architecture," in *2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, June 2015, pp. 1–9.

[13] OpenAirInterface$^{\mathrm{TM}}$ Software Alliance, "OpenAirInterface," accessed Apr. 15, 2019. [Online]. Available: https://www.openairinterface.org/

[14] A. Achtzehn, J. Riihijärvi, and P. Mähönen, "Improving accuracy for tvws geolocation databases: Results from measurement-driven estimation approaches," in *2014 IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, April 2014, pp. 392–403.

[15] S. Koya, K. Masayuki, I. Kei, and F. Takeo, "Measurement-Based Spectrum Database for Flexible Spectrum Management," *IEICE Transactions on Communications*, vol. E98.B, no. 10, pp. 2004–2013, 2015.