Delayed Offloading Zone Associations using Cloud Cooperated Heterogeneous Networks

1,2Ehab Mahmoud Mohamed, 1Kei Sakaguchi, and 1Seiichi Sampei
1Graduate School of Engineering, Osaka University, 2Electrical Engineering Dept., Aswan University.
Email: ehab@wireless.comm.eng.osaka.ac-u.ac.jp, {sakaguchi, sampei}@comm.eng.osaka-u.ac.jp

Abstract—Delayed offloading of mobile user delay tolerant traffic, through other networks such as WiFi networks or millimeter wave (mm-w) gates, is a promising solution for dealing with the ever-growing problem of wireless cellular network capacity. In this paper, we investigate the problem of how we can optimally associate a user with delayed traffic to a nearby offloading zone, i.e., to which nearby offloading zone a user with delayed traffic should be associated. Towards that, we propose an adaptive scheme for offloading zone association, including a weighted proportional fairness (WPF) online algorithm for optimal zone selection. The proposed scheme is a network-based, which maximizes the total system offloading and the users offloading experiences (OFEs). To efficiently implement the proposed scheme, we also suggest a cloud cooperated heterogeneous network (CC-HetNet) to hold the delayed offloading process in a cloud cooperated manner. In this CC-HetNet, the offloading zones and the Macro Base station (BS) are linked to the Centralized Radio Access Network (C-RAN) via ultra-high speed backhaul links. The proposed scheme of offloading zone association will be implemented in the C-RAN as an enhanced Access Network Discovery and Selection Function for delayed offloading networks.

I. INTRODUCTION

Due to the huge proliferation of smart phone and tablet users, the cellular network capacity comes at the bottleneck. Many studies claim that by the end of the year 2014, a broadband mobile user will consume an average traffic of 7 GB per month with an expected exponential increase of 2-fold yearly [1] [2] [3]. Consequently, many researchers pay the attention to how we can increase the current cellular network capacity to fit the expected huge increase in mobile data traffic [1] [2], [4]-[6].

Nowadays, many of the generated mobile traffic are delay tolerable, i.e., the traffic can be delayed for some predefined time without affecting user satisfactions such as movie/game downloading, mobile backups... etc. Accordingly, many researchers are investigating the delayed offloading concept as a hopeful solution to the limited capacity of the cellular networks. In this scheme, each data transfer is associated with a fixed deadline, and the data transfer is resumed whenever the user is being under the coverage of an offloading zone (WiFi hotspot or Millimeter Wave gate) until the transfer is completed. If the transfer does not complete within its deadline, cellular network completes the transfer [2] [5] [7] [8]. The researches proved the significance of delayed offloading in reducing the demand of cellular network transmission through field measurements and statistical analyses. But, they have not shown a complete network paradigm needed to perform the Access Network Discover and Selection Function (ANDSF) [9] to efficiently include the delayed offloading mechanism. The ANDSF mechanism is required for discovering the access networks (offloading zones) located close to the mobile device. At the best of our knowledge, in all current delayed offloading studies, the ANDSF is done autonomously by the User Equipment (UE) at the expense of the UE energy consumption, or by applying the conventional network-based ANDSF with minimal control over access networks [2] [5] [7] [8]. In the conventional ANDSF, the network only discovers the available offloading zones around the UE based upon its location. Then, it sends the list of these available hotspots along with their accessing policies back to the UE over a logical interface. Therefore, the user can select one of these hotspots to connect with, which is usually the nearest hotspot available for him. In both association cases, the user himself, who selects the offloading zone, with no guarantee that the selected zone has a remaining offloading capacity relevant to his delayed traffic. A user may select the nearest WiFi hotspot to offload his delayed traffic, which is heavily loaded, while there is another lightly loaded WiFi hotspot in that location. Thus, the maximization of the total network offloading is not guaranteed with no maximization of the users offloading experiences (OFEs).

In this paper, we investigate the problem of network-based zone selection for associating a user with delayed traffic to the optimal offloading zone from his nearby discovered zones. In this regard, we propose an adaptive process for zone association, including a network-wide weighted proportional fairness (WPF) online algorithm for selecting the optimal associating zone. In this procedure, the optimal offloading zone selection is adaptively changed not simply based upon the changes on the user’s trajectory, but also based upon the changes on his delayed traffic. Hence, it copes with the delayed offloading concept. In the proposed scheme, the optimal zone selection is managed through a network-wide WPF online algorithm, which assigns a newly arrived user, a user seeking for offloading zone association, to an offloading zone that most improves the proportional fairness objective without changing existing users association. The WPF objective is to maximize the total system offloading and the users OFEs by considering the users’ delayed traffic. Increasing users OFEs will increase their satisfaction, because they postpone a file transmission until meeting a high-capacity offloading zone at the aim of reducing their UEs energy consumption. Beside the network-based optimal zone selection, we also investigate the proposed adaptive zone association using a network-based random zone selection.

To efficiently carry out the proposed scheme of network-based zone association, we propose a cloud cooperated heterogeneous network (CC-HetNet), in which the proposed scheme can be implemented as an enhanced ANDSF (eANDSF) for delay tolerant networks (DTN). In the proposed CC-HetNet, the offloading zones and the Macro BS are connected in a centralized manner to the Centralized Radio Access Network (C-RAN) via ultra-high speed backhaul links. Utilizing the concept of control/user (C/U) plane splitting [6], the control signals, such as the list of discovered nearby zones with the optimally/randomly selected one, are sent to the UE through the wide coverage of the Macro BS.
In addition, the user’s delayed traffic is mainly offloaded by the high-capacity offloading zones. This CC-HetNet enables us to fully monitor and control the delayed offloading process. In which, the user generates a list of delay tolerant files with different sizes and deadlines to be uploaded/downloaded. This file list will be registered at the UE in the form of delayed files table. The C-RAN always keeps track of this files table and its updates through the Macro-BS signaling. After estimating the user current location, the C-RAN discovers his nearby offloading zones. Then, it selects an optimal/random offloading zone for his association using the proposed WPF algorithm/uniform random generator, respectively. As an application, we apply the proposed scheme on a CC-HetNet consisting of the ultra-high capacity mm-w gates proposed in [8].

The rest of this paper is organized as follows; Section II presents the details of the proposed delayed offloading CC-HetNet using the mm-w gates. The proposed adaptive process for zone association is given in Sect. III. Section IV gives the simulation analysis followed by the conclusion in Sect. V.

II. THE DELAYED OFFLOADING CC-HETNET ARCHITECTURE AND PROTOCOL

Figure 1 shows the detailed structure of the proposed CC-HetNet including the offloading zones. In this figure, we use the mm-w gates proposed in [8] as ultra-high capacity offloading zones. Although, in this paper, we use mm-w gates as offloading zones, the concept is general and can be applied on whatever type of offloading zones such as WiFi hotspots. Mm-w gate is proposed by the authors in [8] as an ultra-high capacity mm-w gate [10]. These mm-w APs are controlled by a local coordinator inside the gate [8] [11]. The gate coordinator is responsible for the Radio Resource Management (RRM) inside the gate, such as mm-w APs associations, re-associations, joint beamforming and user scheduling… etc [8]. Hence, it is responsible for maximizing the gate offloading efficiency with maintaining the long term offloading fairness among the users inside the gate [8]. Besides, it works as a Gateway to connect the mm-w gate to the cellular network. In Fig. 1, mm-w gate 3 consists of 4 mm-w APs controlled by a local coordinator installed inside the gate. In the proposed CC-HetNet, the offloading zones (local coordinators in case of mm-w gates) and the Macro BS are linked to the C-RAN via ultra-high speed backhaul links.

The proposed CC-HetNet protocol is shown in Fig. 2. The C-RAN is responsible for controlling the delayed offloading process in the proposed CC-HetNet. Accordingly, the C-RAN sends/picks up the signaling information to/from the UEs through the Macro-BS control channel, respectively. As well, the C-RAN performs the UE location estimation. This can be done with good accuracy and low power consumption by using the Observed Time Difference Of Arrival (OTDOA) location estimation technique or by performing location estimation using the small cells deployed in the HetNet [13]. In addition, observing/recording the details of the user’s daily habits can be used to relax the requirements of instantaneous UE location estimation. The proposed eANDSF is also performed by the C-RAN. Hence, based upon the estimated UE location and the positions of the registered offloading zones, the C-RAN discovers the available offloading zones around the UE. From these discovered zones, the C-RAN selects an optimal/random offloading zone for the user association. The optimal zone selection is performed through the proposed network-wide WPF online algorithm, in which the users are linked up with the offloading zones that maximize the total zones offloaded bytes and the users OEFs. Also, the random zone selection is done using a uniform random generator. The C-RAN sends the list of discovered zones including the optimally/randomly selected one back to the UE through the Macro BS signaling with an information about the estimated zones reaching time (ZRT) and how to reach. Furthermore, the C-RAN switches ON/OFF the UE WiFi or mm-w module before it enters/after it leaves the offloading zone, respectively. This switching ON/OFF functionality greatly reduces the UE energy consumption. The C-RAN always has the updated version of the UEs delayed files tables collected by the Macro-BS signaling. Hence, it can efficiently select an optimal offloading zone for the UE. In addition, whenever some mobile users reach at a certain offloading zone, it passes their files tables to this zone, to be employed for data prefetching, RRM, and files offloading [11] [13].
Fig. 3. The proposed adaptive process for zone association to a user with delayed traffic.

On the UE side, to effectively control the delayed offloading process, each UE has an offloading organizer (scheduler) software application that coordinates the file offloading process. This software can be easily implemented using typical smart phone platforms. As the UE generates a delay tolerant file such as movie downloading or mobile backup, it will be registered in the software, including the file necessary information such as its title, total size and deadline. The file deadline is typically assigned by the user or an application program. In addition, the user can see the information of the discovered zones, including the optimally/randomly selected one, e.g., the Service Set Identifier (SSID), the estimated ZRT and how to reach using a Google map like interface built into the organizer software, as illustrated in Fig.1. The zones information is sent by the C-RAN to the UE organizer software through the Macro BS signaling. Hence, the user can easily join the designated zone within the estimated ZRT. Sometimes, the user selects an offloading zone other than the one selected by the proposed eANDSF, for example the user urgently wants to go to a certain place, e.g., train station. In such cases, the offloading software will inform the C-RAN about the user selection to take into account when selecting the optimal offloading zones for the other users. The delayed files are scheduled for transmission in shortest remaining time first (SRTF).

Figure 2 gives the proposed protocol that organizes the operation between the basic elements of the proposed delayed offloading CC-HetNet, i.e., the UE (organizer software), the offloading zone, the Macro BS and the C-RAN (eANDSF).

III. THE PROPOSED ADAPTIVE OFFLOADING ZONE ASSOCIATION

In this section, we propose the adaptive process needed to associate a user with delayed traffic to an optimal/random offloading zone from his nearby discovered zones. This algorithm will be implemented in the C-RAN as an eANDSF for delayed offloading networks. Figure 3 shows the details of the proposed adaptive process. In this process, as the user generates a new delayed file with a new assigned deadline, or as the user changes his position, the process of optimal/random offloading zone selection is re-initiated. Using the eANDSF, the optimal/random offloading zone selection is executed after the C-RAN discovers the available offloading zones around the UE based upon its estimated location. If the user selects an offloading zone, other than the network-selected one, the UE organizer software will inform the C-RAN about his selection. Hence, the C-RAN takes his choice into account when it selects the optimal offloading zones for the other users. If the user moves away from the network-selected zone, and he does not select any offloading zone by himself, the process of optimal/random zone selection will be re-started.

A. The Proposed WPF Algorithm for Optimal Zone Selection

In selecting an optimal offloading zone for a user with delayed traffic, from his nearby discovered zones, the C-RAN should maximize the total zones offloaded bytes and the users OFE, which is defined as:

\[
OFE = \frac{\sum_{\text{User}} \text{bytes successfully offloaded by the offloading zones}}{\text{User total delayed bytes}}
\]  

(1)

Suppose that some mobile users having different velocities arrive at the area of different capacity offloading zones. Each user has a number of delayed files with different remaining sizes and delay times. At the zone selection time \( t \), let \( l_k(t) \) indicates the total number of delayed bytes indexed in the organizer software of user \( k \). Also, each user has different estimated reaching times to the available offloading zones \( ZRT_{zk} \), where \( ZRT_{zk} \) is the estimated time required by user \( k \) to reach an offloading zone \( z \) in seconds. \( ZRT_{zk} \) mainly depends upon the zone location, the estimated UE location and the estimated user speed. Based upon \( ZRT_{zk} \) and the remaining delay times of user \( k \) delayed files, the C-RAN can expect the load in bytes of user \( k \) on offloading zone \( z \), \( l_{zk}(t) \). \( l_{zk}(t) \) is different from zone to zone for the same user. \( l_{zk}(t) \) can be estimated as:

\[
l_{zk}(t) = l_k(t) - \sum_{f \in F_{zk}(t)} \left( \pi_{zk}^f(t) \right),
\]

(2)

\[
\pi_{zk}^f(t) = \mu(ZRT_{zk}(t) - \Gamma_k^f(t)), \quad \Gamma_k^f(t) < ZRT_{zk}(t),
\]

(3)

where \( F_{zk}(t) \) is the total number of user \( k \) delayed files expected to have expired deadlines before user \( k \) reaches the offloading zone \( z \). \( F_{zk}(t) \) can be easily estimated by the C-RAN via estimating \( ZRT_{zk}(t) \) and knowing the files remaining delay times thanks to the C/U splitting concept. \( \pi_{zk}^f(t) \) is the total number of delayed bytes that will be transmitted from the expired deadline file \( f \) through the Macro BS before user \( k \) reaches the offloading zone \( z \). \( \mu \) is the Macro BS transmission rate in bytes per second. \( \Gamma_k^f(t) \) is the remaining delay time of file \( f \) in second.

Based upon these expectations, the C-RAN should select an optimal offloading zone from user \( k \) nearby zones for his association. This optimal zone selection should maximize the total zones offloaded bytes and the users OFE. The trivial solution is to minimize the amount of delayed bytes transmitted by the Macro-BS by associating all users to their nearest zones with the smallest \( ZRT_{zk}(t) \). The nearest zone association is used by the conventional ANDSF, in which the user always joins the nearest zone available for him. Although this solution highly decreases the amount of delayed bytes transferred by the Macro-BS when the user moves toward the selected zone, it does not care either the zones’ loads at the user arrival or their offloading capacities that highly affect the total offloading efficiency. The user may join the nearest offloading zone, with the aim of reducing the number of delayed bytes transmitted by the Macro-BS, but unfortunately this offloading zone has a remaining capacity un-relevant to the size of his delayed traffic. As a result, the total zones offloaded bytes will be greatly reduced with a high decrease in the users OFE.

In this paper, we propose a network-wide WPF algorithm to distribute the users’ delayed traffic among the offloading zones...
with the goal of maximizing the total zones’ offloaded bytes and the users OFEs. A zone association algorithm is weighted proportional fair, if and only if the following inequality is satisfied for any feasible algorithm [14]:

\[
\sum_{z=1}^{Z} w_z \frac{L_x - L_z}{L_z} \leq 0, \tag{4}
\]

where \( Z \) indicates the total number of nearby zones a user \( k \) can reach. \( L_x \) is the average load assigned to an offloading zone \( z \) by any feasible zone association algorithm, and \( L_z \) is the zone’s average load assigned by a WPF algorithm. \( w_z \) is the average offloading capacity of zone \( z \) normalized to the lowest average offloading capacity zone, which is different from zone to zone, and it is directly related to the type of AP, the number of cooperating APs, and the RRM inside zone [8].

The WPF in (4) can be expressed as the following maximization problem [14]:

\[
\max \left( \sum_{z=1}^{Z} w_z \log(L_z) \right). \tag{5}
\]

Hence, in addition to maximizing the total average loads assigned to the zones, zones are assigned loads with respect to their offloading capacities, i.e., higher capacity zones should be assigned more loads, and vice versa. For simplification, let the averaging process be executed through a weighted low pass time window, therefore the optimization problem can be re-written as [14]:

\[
\max \left( \sum_{z=1}^{Z} w_z \log(L_z(t)) \right), \tag{6}
\]

where \( L_z(t) \) can be calculated as:

\[
L_z(t) = \left(1 - \alpha_t\right)L_z(t-1) + \frac{1}{\alpha_t} l_{zk}(t)a_{zk}(t), \tag{7}
\]

where \( L_z(t) \) is the expected average load assigned to zone \( z \) if user \( k \) will join it at the time of zone selection \( t \), i.e., after \( ZRT_{zk}(t) \). \( L_z(t-1) \) is the expected average load inside the zone just before user \( k \) joining. \( N \) is the averaging low pass filter window size. \( a_{zk}(t) = 1 \) if zone \( z \) is selected for associating user \( k \) and 0 otherwise. User \( k \) is assumed to be allocated to only one offloading zone, i.e. \( \sum_{z=1}^{Z} a_{zk}(t) = 1 \), which is the constraint of the maximization problem in (6) and (7), and \( l_{zk}(t) \) is given in (2).

From Taylor series expansion:

\[
\log(L_z(t)) \approx \log(L_z(t-1)) + \frac{1}{L_z(t-1)} \left( L_z(t)-L_z(t-1) \right), \tag{8}
\]

Hence, from (2) and (7), (8) can be formulated as:

\[
\log(L_z(t)) \approx \log(L_z(t-1)) + \frac{1}{N} \left( \frac{L_z(t)-\sum_{z=1}^{Z} \frac{\pi f_{zk}(t)}{L_z(t-1)} a_{zk}(t) - 1}{L_z(t-1)} \right). \tag{9}
\]

Thus, the solution of the optimization problem, given in (5), becomes:

\[
z^*(t) = \arg \max_{z} \left( w_z \frac{l_{zk}(t)-\sum_{z=1}^{Z} \frac{\pi f_{zk}(t)}{L_z(t-1)}}{L_z(t-1)} \right). \tag{10}
\]

Therefore, at the zone selection time \( t \); if there are two available offloading zones for user \( k \) with equal expected average loads at the user arrival, \( L_z(t-1) = L_z(t-1) \), and equal offloading capacities \( w_1 = w_2 \), the closer zone to the user (smaller \( ZRT_{zk}(t) \)) will be selected by the C-RAN for the user association. This will highly reduce the number of delayed bytes transmitted by the Macro-BS in the user way toward the offloading zone. At the same time, if the two zones have the same offloading capacities \( w_1 = w_2 \), and user \( k \) needs the same time to reach them, \( ZRT_{1k}(t) = ZRT_{2k}(t) \), the zone with the lower average load at the user arrival will be selected for his association. Finally, if they have the same average loads at the user arrival, \( L_1(t-1) = L_2(t-1) \), and user \( k \) needs the same time to reach them, \( ZRT_{1k}(t) = ZRT_{2k}(t) \), the C-RAN selects the higher offloading capacity zone for the user association. Thus, the proposed WPF algorithm for optimal zone selection always maximizes the total zones’ offloaded bytes. In addition, maximizing users OFE is guaranteed using the proposed algorithm. This is because the proposed algorithm always selects an offloading zone that maximizes \( l_{zk}(t) \) by decreasing the amount of delayed bytes transmitted by the Macro-BS. As well, the selected zone has a remaining capacity, at the user arrival, relevant to the user’s delayed traffic size \( l_{zk}(t) \).

B. The Proposed Online-Algorithm for Optimal Zone Selection

Based upon the proposed WPF criterion, the following online algorithm is used by the C-RAN to select the optimal offloading zone for associating a user \( k \) form his nearby discovered zones. In this algorithm, \( ZRT_{\max} \) is the zone reaching time of the farthest offloading zone. \( ZRT_{\max} \) is a design parameter used by the eANDSF for discovering the list of user \( k \) nearby zones.

\[
\begin{align*}
\text{Algorithm Online optimal zone selection} \\
& t \text{ is the time of zone selection for user } k, \text{ and } t-1 \text{ is the time immediately before the selection decision.} \\
\text{Input: User } k \text{ delayed traffic size } l_{zk}(t), \text{ the estimated UE location and velocity, files remaining deadlines } l_{zk}(t), \forall z \in \{ZRT_{zk}(t) \leq ZRT_{\max}\}, \text{ zones locations, zones } w_z \text{ values.} \\
& 1. \text{Estimate } ZRT_{zk}(t) \\
& 2. \text{Estimate } F_{zk}(t), \sum_{f=1}^{F_k} (\pi f_{zk}(t)) \text{ and } l_{zk}(t). \\
& 3. \text{Estimate } L_{zk}(t-1). \\
& 4. \text{Select an offloading zone as:} \\
& z^*(t) = \arg \max_{z} \left( w_z \left( \frac{l_{zk}(t)}{L_{zk}(t-1)} \right) \right)
\end{align*}
\]

Fig. 4. The proposed online-algorithm for optimal offloading zone selection.

IV. SIMULATION ANALYSIS

In this section, we give the simulation analysis that prove the effectiveness of the proposed eANDSF over the conventional ANDSF using mm-w gates as ultra-high capacity offloading zones.

A. Simulation Scenario and Simulation Parameters

In the simulation scenario, we assume a number of randomly distributed UEs inside the simulation area of 2km x 2km. Likewise, a number of mm-w gates with different average offloading capacities are randomly distributed in the area. Each UE has a random velocity in the range of \([5, 50] \text{ km/hr.} \) Hence, the maximum \( ZRT \) \( ZRT_{\max} \) will be in the range of 30 minutes (our definition of the farthest nearby zone). Although we assume ideal UE location estimation, practical UE location estimation can be easily considered by introducing some errors in the estimated \( ZRT_{zk}(t) \) resulted from the errors in the UE location estimation. Also, we assume that each user has a random number of delayed files with different sizes and remaining times to their deadlines. This delayed traffic can be a remaining traffic after the user leaves an offloading zone (mm-w gate), or it can be initially generated by the user for the first time. The total size of a user delayed traffic is generated using a uniform random distribution in the range of \([1, 15] \text{ GB.}\) Without loss of generality and for the purpose of fair comparisons between the compared schemes, we assume that all users do not generate any delayed files until reaching their associating gates. In addition, the users always move toward their selected gates; although the proposed adaptive process can handle any case. For each delayed
file, a randomly generated delay time is assigned to it in the range of [0, 90] minutes. The zero deadline file is a file with an expired deadline. The files with expired deadlines before the user reaches the selected gate will start to be transmitted using the Macro BS. According to the study given by the authors in [8], we assume that the user’s average stay time inside the gate coverage is 30 Sec. After the user leaves the gate, the Macro BS switches OFF the UE mm-w module and the entire user’s remaining traffic will be transmitted using the Macro BS. Table 1 summarizes the used simulation parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation area.</td>
<td>2km x 2km</td>
</tr>
<tr>
<td>Num. of UEs.</td>
<td>500 ~ 4000</td>
</tr>
<tr>
<td>Num. of mm-w gates.</td>
<td>2 ~ 30</td>
</tr>
<tr>
<td>Gate average offloading capacity</td>
<td></td>
</tr>
<tr>
<td>( C_g ) as proposed in [8].</td>
<td>( C_g )=[6.5 (1-AP) 12 (2-APs) 18 (3-APs) 22 (4-APs)] Gbps. ( w_g )=[1 1.85 2.77 3.4].</td>
</tr>
<tr>
<td>Files remaining delay time</td>
<td>Uniform random distribution in the range of [0, 90] minutes.</td>
</tr>
<tr>
<td>(deadline) model.</td>
<td></td>
</tr>
<tr>
<td>The user’s average stay time</td>
<td>30 Sec.</td>
</tr>
</tbody>
</table>

**B. The Compared Association Schemes**

In simulation comparisons, we compare the performance of three zone association algorithms. The first one is the conventional ANDSF with the user selects the shortest ZRT zone. In which, the network only discovers the offloading gates around the user, and the user selects the shortest time reachable gate to associate with. Conventional ANDSF is the currently proposed scheme for offloading zones discovery and selection for delayed offloading networks [9]. The second scheme is the proposed eANDSF, but the network selects a random zone from the user nearby discovered zones for his association. The random zone selection is done using a uniform random generator. The third scheme is the proposed eANDSF using the proposed WPF online algorithm for optimal zone selection.

**C. Performance Metrics**

In the analyses, we concern in measuring the followings:

1) Total Gates Offloaded Bytes

The total gates offloaded bytes can be calculated as follows:

\[
G_{BOF} = \sum_{g=1}^{G} \sum_{t=1}^{\text{finl}} \min(l_g(t), C_g(t)), \quad (11)
\]

\[
l_g(t) = \sum_{k=1}^{K_g(t)} l_{gk}(t), \quad (12)
\]

where, int is the simulation start time, and finl is the simulation end time. For calculating the total bytes offloaded by a gate \( g \), at every second \( t \), we count the total number of users inside the gate \( K_g(t) \). \( K_g(t) \) is used by the gate coordinator to assign the number of cooperative mm-w APs \( N_{CAP} \) required for joint transmissions with total gate offloading capacity of \( C_g(t) \), as shown in Table 1 [8].

\[
N_{CAP} = \begin{cases} K_g(t) & K_g(t) < N_{AP} \\ \left\lceil \frac{K_g(t)}{N_{AP}} \right\rceil & K_g(t) \geq N_{AP} \end{cases}, \quad (13)
\]

where \( N_{AP} \) is the total number of mm-w APs inside the gate. Then, the total traffic load inside the gate \( l_g(t) \) is calculated using (12) and compared with \( C_g(t) \) using (11) to find out the total number of bytes a gate \( g \) can offload at every second \( t \).

2) Average User OFE

The average user OFE can be calculated as:

\[
\frac{1}{K} \sum_{k=1}^{K} \text{OFE}(k), \quad (14)
\]

where \( K \) is the total number of users, and \( \text{OFE}(k) \) is the offloading experience of user \( k \) as defined in (1).

**V. SIMULATION RESULTS**

Figure 5 shows the total gates offloaded bytes in Tera Byte [TB] using a fixed number of users (2000 users) and different number of gates (2~30 gates); while Fig. 6 shows the total gates offloaded bytes using a fixed number of gates (8 gates) and different number of users (500~4000 users). These figures prove the high significance of the proposed eANDSF in maximizing the total system offloaded bytes compared to the conventional ANDSF. The proposed eANDSF with network-based random zone selection has an increase in performance of 1.2 compared to the conventional ANDSF with the nearest zone association. Also, the eANDSF performance reaches an increase in performance of 1.6 compared to the conventional scheme if we use the proposed WPF online algorithm for optimal zone selection. The eANDSF with random zone association has a higher performance than the conventional ANDSF because the network randomly allocates the users’ delayed loads to the offloading zones, which results in less congested zones compared to the conventional scheme, in which all users are associated to the nearest zones.

Figures 7 and 8 show the average user OFE using different number of gates with a fixed number of users, and using different number of users with a fixed number of gates, respectively. As it is clearly appeared, the proposed eANDSF schemes have higher average user OFE performance than the conventional scheme. This is because, in the proposed eANDSF, the network distributes the users’ delayed loads randomly or optimally using the WPF online algorithm among the offloading zones. Consequently, some load balance is guaranteed among the offloading zones, which results in higher total gates offloaded bytes and higher users OFE compared to the conventional ANDSF. From these results, the eANDSF with random zone selection is recommended for low complexity delayed offloading networks with sub-optimal offloading performance.
In this paper, as a contribution to embed the delayed offloading to be a component of future cellular networks, we explored the problem of offloading zones discovery and selection. Towards that, we proposed an adaptive scheme for offloading zone association. In this scheme, a novel network-based WPF online algorithm is proposed to select the optimal offloading zone for a user with delayed traffic. This adaptive scheme will be used as enhanced Access Network and Discovery Function (eANDSF) for delay tolerant networks. To efficiently implement the proposed eANDSF, we proposed a cloud cooperated heterogeneous network to control the delayed offloading process in a centralized manner, in which, the offloading zones and the Macro-BS are linked to the C-RAN via high speed backhaul links. The proposed WPF online algorithm enables the C-RAN to select the optimal offloading zone based upon the user’s delayed traffic relative to the nearby zones remaining capacities, at the expected user arrival. We proved that the proposed eANDSF outperforms the conventional ANDSF when applied for delayed offloading networks in terms of the total system offloading and the user offloading experience.

ACKNOWLEDGMENT
This work is partly supported by “Research and development project for expansion of radio spectrum resources” of MIC, Japan.

REFERENCES