IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 30, 2017, accepted January 16, 2018, date of publication February 2, 2018, date of current version April 4, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2801386

Delay-Aware LTE WLAN Aggregation in Heterogeneous Wireless Network

BIN LIU, (Student Member, IEEE), QI ZHU, AND HONGBO ZHU
Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Corresponding author: Hongbo Zhu (zhuhb@njupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61427801, Grant 61571234, and Grant 61631020, and in part by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant KYLX16_0659.

ABSTRACT The scarcity of spectrum turns the next generation wireless network increasingly heterogeneous. Recently, much attention has been focused on exploiting the unlicensed spectrum to leverage traffic burden from the cellular network. Specified in 3GPP Release 13, long term evolution (LTE) wireless local area networks (WLAN) aggregation (LWA) promises an effective approach for the LTE WLAN interworking and unlicensed spectrum utilization. The LWA could improve the data rate by the licensed and unlicensed carriers aggregation. In contrast to the previous studies, where the LWA is always chosen to boost the peak data rate without considering the user payment, we jointly consider the LWA with Wi-Fi offloading in this paper, aiming to strike the balance between user payment and quality of service (QoS) requirement. We formulate the multi-slot modes selection as a finite-horizon Markov decision problem and propose a Delay-Aware LTE WLAN Aggregation (DLWA) algorithm to obtain the optimal transmission modes strategy. Due to the computational complexity of sequential decision problem, we exhibit the threshold structure of LTE usage and develop a low complexity algorithm based on this structure. Moreover, imperfection in LWA backhaul is considered and analyzed in this paper. Simulation results show the DLWA algorithm and threshold-based DLWA algorithm could guarantee the QoS requirement with lower user payment compared with current LTE WLAN interworking schemes.

INDEX TERMS Long term evolution (LTE) wireless local area networks (WLAN) aggregation, unlicensed spectrum, WiFi offloading, heterogeneous network, mode selection.

I. INTRODUCTION

With the proliferation of smart devices such as smartphones and tablets, cellular networks are facing an exponential growth of mobile data traffic. According to Cisco's forecast, global mobile data traffic is expected to grow to 49 exabytes per month by 2021, a sevenfold increase over 2016 [2]. Constrained by limited bandwidth, the cellular network capacity, however, cannot keep up with the explosive data growth [3]. One promising solution is to utilize unlicensed spectrum. Owning to the existing WiFi deployment, WiFi offloading is a straightforward method to leverage the traffic load. It has been shown in [4] that the WiFi offloading can leverage more than 65% traffic from the cellular network.

WiFi offloading does not allow packet flow aggregation over long term evolution (LTE) and wireless local area networks (WLAN) access [5]. The single connection mechanism of offloading may lead to a low data rate, and thus degrades the QoS performance. To this end, in March 2016, LTE WLAN Aggregation (LWA) solution was formally approved by 3GPP RAN Plenary in Release 13 [6]. Unlike WiFi offloading, the data packets of a connection are split over both the cellular and WiFi networks simultaneously with LWA, which attracts much attention. On 19 August 2016, Singapore M1 with Nokia announced Singapore's first commercial LWA heterogeneous network rollout, and expected the LWA could increase the peak download speed to more than 1 Gbps by 2017 [7].

A. RELATED WORKS

Early studies of WiFi offloading mainly focused on the offloading efficiency, which is defined as the ratio of the offloaded data to the total data amount. Lee *et al.* [4] studied the WiFi offloading performance through an experiment in Seoul, and proposed an on-the-spot offloading scheme, which offloads user's data to the WiFi network whenever available. An analytical model was developed in [8] to analyze the offloading efficiency. These studies can be regarded as simple opportunistic WiFi offloading attempts, in which the offloading decision only depends on the WiFi availability. Another line of studies focused on the delayed offloading.

As the deadline is set to be a large value, the user has more opportunities of accessing WiFi networks, which could increase the amount of offloaded data. For example, Balasubramanian *et al.* [9] demonstrated that a larger portion of cellular traffic can be offloaded to WiFi with the delayed offloading scheme. Deng and Hou [10] studied the capacity of delayed offloading without prior knowledge of users' mobility patterns, and proposed online scheduling policy to maximize the amount of offloaded data. Wang and Wu [11] and Cheung and Huang [12] took consideration of both the downloading cost and delay, and maximized the user's satisfaction by dynamically selecting networks within the deadline. The optimal transmission deadline was derived by Ko *et al.* in [13] to save monetary cost while maintaining the outage probability. Lee *et al.* [14] further investigated the economic benefits for both the operators and users in the delayed offloading scheme.

The aforementioned WiFi offloading [8]–[14] does not permit aggregation of packet flows over LTE and WLAN access, which degrades the QoS performance and thus cannot fully utilize the unlicensed spectrum. For instance, in the delayed offloading, the user prefers to access the LTE network when the deadline is tight, rather than to choose the low-rate WiFi network. To this end, 3GPP approved LWA solution in Release 13 to support the access to the cellular and WLAN network simultaneously by carrier aggregation. Previous LWA studies mainly focused on the prototype and architecture design [5]. Zhu *et al.* [15] experimentally verified the feasibility of licensed and unlicensed carriers aggregation. Ohta *et al.* [16] developed the layer 2 (L2) structure for LWA to achieve the compatibility with WLAN. Further, the load balancing and user assignment solutions for LWA were investigated in [17].

The aforementioned studies [5]–[17] nevertheless did not consider the user payment in LWA decision, and neglected to exploit advantages of WiFi offloading to reduce user payment. Most of the studies adopted LWA whenever possible, even when WiFi offloading alone can meet the requirement, which may lead to a higher user payment due to LTE data usage in aggregation. Clearly, WiFi offloading could reduce the user payment at the cost of poor QoS performance. On the contrary, the LWA could guarantee QoS requirement but cost more. Therefore, there is a tradeoff between the QoS and user payment, and thus the LWA and WiFi offloading should be jointly considered to obtain a better tradeoff compared with other LTE and WLAN interworking schemes. In addition, the aforementioned studies are based on the assumption that the aggregation is ideal with the perfect backhaul, which is unrealistic for the carrier aggregation. Singh *et al.* [18] demonstrated the rate loss of no-ideal backhaul in LWA. The packet delay and re-ordering latency were considered in the carrier aggregation between licensed and unlicensed band [19]. Through the no-ideal backhaul of LWA is observed, few works further model the imperfection and analyze its impacts on aggregation decision.

B. OUR CONTRIBUTIONS

Motivated by the aforementioned observations, in this paper, we jointly consider the LWA with WiFi offloading to strike the balance between user payment and QoS. Here, the QoS requirement is characterized by the completion probability of data transmission within the deadline. Based on the WiFi availability and transmission deadline, four transmission modes can be dynamically chosen for each user, which include 1) keep idle (i.e., waiting for next slot), 2) use LTE directly, 3) operate WiFi offloading, and 4) perform LWA. In this paper, the approach in [12] is extended to tackle with the multi-slot transmission modes selection problem, which can be formulated as a finite-horizon sequential Markov process [20]. To obtain the optimal modes strategy, we propose Delay-Aware LTE WLAN Aggregation (DLWA) algorithm based on dynamic programming (DP). However, the sequential decision problem is computationally intractable. To this end, threshold structure of LTE usage in remaining data size and time is revealed based on monotonicity of optimal selection strategy. The threshold-based DLWA (TB-DLWA) algorithm is proposed with much lower computational complexity, whose performance is close to DLWA. The main contributions of our work are summarized as follows:

- *Delay-Aware LTE WLAN Aggregation algorithm*: To our best knowledge, it is the first paper jointly consider LWA with WiFi offloading for unlicensed spectrum utilization. By viewing the multi-slot modes selection as the finite-horizon Markov decision problem, we propose the Delay-Aware LTE WLAN Aggregation algorithm to obtain the optimal selection strategy, which could harvest the tradeoff between user payment and QoS.
- *Non-ideal Backhaul in LWA*: We take the imperfection in LWA backhaul into the mode selection and analyze the impacts on aggregation decision. We reveal that LWA is chosen only when the aggregation gain is larger than the minimum value, which is determined by the average LTE and WiFi rate.
- *Low-complexity threshold-based DLWA algorithm*: With a convex penalty, we prove the threshold structure of LTE usage in remaining data and time. Based on the threshold structure of LTE usage, we develop the low-complexity threshold-based DLWA algorithm.

The rest of the paper is organized as follows. The scenario for LWA is described in Section II. In Section III, we formulate the modes selection problem. DLWA algorithm is proposed in Section IV. We exhibit the threshold structure of LTE usage, and further develop the threshold-based DLWA algorithm in Section V. Performance evaluation is conducted in Section VI, and conclusions are drawn in Section VII.

II. SYSTEM MODEL

We consider bearer-split framework in LWA deployment [21], as illustrated in Fig. 1, where the LTE eNodeB (eNB) acts as the scheduler to split the flow, and the WLAN access points (APs) are the boosters.

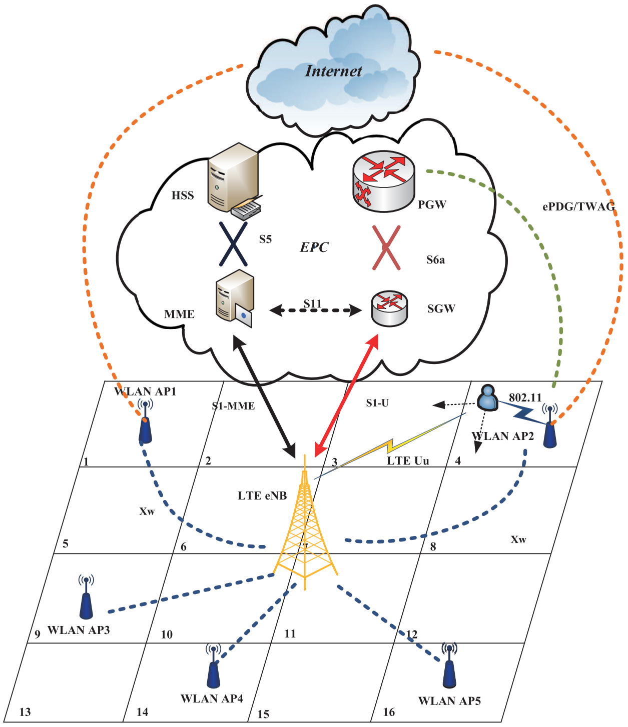


FIGURE 1. The scenario for non-collocated LTE WLAN Aggregation.

The non-collocated connection between eNB and WLAN APs is by the standardized interface X_w. In this paper, we consider the downlink aggregation sits at the Packet Data Convergence Protocol (PDCP) layer, where the eNB configures the transmission mode for the user. We combine the LWA together with WLAN offloading, which allows APs to be directly connected to the Core network via ePDG/TWAG.

In the paper, the large-scale fading channel is considered in LTE network. In particular, the large-scale fading coefficient of the LTE eNB to the k-th user is given by $h_k = d_k^{-\alpha_0}$, where d_k is the distance between the k-th and the LTE eNB, and α_0 is the path loss exponent. Thus, the achievable rate for the k-th user is given by

$$R(k) = B\log_2\left(1 + \frac{P|h_k|^2}{N_0}\right), \tag{1}$$

where B is the bandwidth of LTE subcarrier, P represents the transmitting power of the LTE eNB, and N_0 is the power of additive Gaussian noise.

The user location is indexed by $\alpha \in \mathcal{A} = \{1, 2, \ldots, A\}$, where A is the maximum number of possible locations. The whole scenario is covered by LTE, while WiFi is available only in these specific areas with APs. The locations set $\mathcal{A}$ is divided into two subsets $\mathcal{A}^{(0)}$ and $\mathcal{A}^{(1)}$ according to where WiFi is not and is available respectively, namely $\mathcal{A}^{(0)} \subseteq \mathcal{A}$, $\mathcal{A}^{(1)} \subseteq \mathcal{A}$, and $\mathcal{A}^{(1)} = \mathcal{A}\backslash\mathcal{A}^{(0)}$, for instance, $\mathcal{A}^{(1)} = \{1, 4, 9, 14, 16\}$ as shown in Fig. 1.

Assume that user equipment (UE) initiates an application and needs to deliver S bits of data within time T, for example, the users want to download a video with a size of 750 Mbytes on commuting through the smart device in the 10 minutes. We divided the predetermined deadline T into time slots with normalized size $\Delta t = 1$, and indexed by $t \in \mathcal{T} = \{1, \ldots, T\}$. The problem lies in how to schedule the transmission modes to reduce user payment within the given time. To be specific, we have transmission modes set as $\mathcal{U} = \{0, 1, 2, 3\}$, where $u = 0$ means that the user will remain idle without data transmission, $u = 1$ means that transmitting only through the LTE network, $u = 2$ indicates that offloading data to WiFi, and $u = 3$ denotes that the transmission is conducted via LTE WLAN carrier aggregation method. Apparently, modes $u = 0$ and $u = 1$ are available for all areas, whereas the modes $u = 2$ and $u = 3$ could be selected only when $\alpha \in \mathcal{A}^{(1)}$. Thus, the mode u depends on the location α, and $u \in \mathcal{U}^{(\alpha)} \subseteq \mathcal{U}$, where $\mathcal{U}^{(\alpha)}$ is the candidate modes set at location α:

$$\mathcal{U}^{(\alpha)} = \begin{cases} \{0, 1\}, & \text{if } \alpha \in \mathcal{A}^{(0)}, \\ \{0, 1, 2, 3\}, & \text{if } \alpha \in \mathcal{A}^{(1)}. \end{cases} \tag{2}$$

The usage-based pricing [22], [23] is adopted in this paper, i.e., user payment is proportionate to the amount of data, which has been adopted by most of the telecom operators, such as China Mobile and Verizon Wireless. Given monetary incentives, the users are willing to wait for downloading [24] and assign the deadline T as maximum time for transmission. The incentives come from the free charge in WiFi access, which is widely used in [11]–[13].

Moreover, we take non-ideal backhaul into the LWA decision. In particular, the backhaul latency t_b denotes the transmission time loss, including packet reordering latency, synchronization hysteresis, etc. Thus, we define the aggregation gain as the ratio of actual transmission time to the time slot, i.e., $G = \frac{\Delta t - t_b}{\Delta t}$. Aggregation gain G is inverse to backhaul latency t_b, and LWA is less favorable with lower gain G.

In this paper, we formulate mode selection problem for multiple slots, which aims to lower the user payment with QoS guarantee. On the one hand, it is incentive to utilize WLAN resource to reduce user payment. On the other hand, the deadline may compel to use LTE network for higher completion, since the user may not have enough opportunities to WiFi networks. Therefore, the optimal mode selection strategy is needed for the better trade-off between user payment and QoS requirement.

III. PROBLEM FORMULATION

In this section, we formulate the multiple-slots modes selection as finite-horizon Markov sequence decision problem [20]. The user state is described as $e = (s, \alpha)$, where $s \in \mathcal{S}$ denotes the remaining data to be delivered, and $\alpha \in \mathcal{A}$ is the location index [12]. $R(\alpha, u)$ is the data rate user achieved at location α with mode u, especially, $R(\alpha, 0) = 0$ for idle mode $u = 0$. $\gamma(\alpha, u)$ represents per unit data price for selecting mode u at location α, particularly $\gamma(\alpha, 0) = 0$, $\forall\alpha \in \mathcal{A}$ for the idle mode. The non-ideal backhaul latency is $t_b(\alpha, u)$, which is nonzero only if $u = 3$. Hence, the user payment function for adopting mode $u \in \mathcal{U}^{(\alpha)}$ at slot $t \in \mathcal{T}$ is expressed as:

$$\begin{aligned} m_t(e,u) &= m_t(s,\alpha,u) \\ &= \min\{s, R(\alpha,u)(\Delta t - t_b(\alpha,u))\} \cdot \gamma(\alpha,u) \\ &= \min\{s, R(\alpha,u)G(\alpha,u)\Delta t\} \cdot \gamma(\alpha,u). \end{aligned} \quad (3)$$

For the QoS requirement, the penalty for not finishing the transmission within deadline is defined as [12], [25]:

$$\overline{m}_{T+1}(e) = \overline{m}_{T+1}(s,\alpha) = \chi(s), \quad (4)$$

where $\chi(s)$ decreases with s, $\chi(s) \geq 0$, and $\chi(0) = 0$. The penalty holds from the $T+1$ time slot (i.e., the next slot after the deadline). $\chi(s)$ can be adjusted according to the time sensitivity of different applications.

The transition probability of user state is described as $p(e'|e,u) = p\left((s',\alpha'),u|(s,\alpha),u\right)$ [25], which denotes the probability of transferring from state $e = (s,\alpha)$ to $e' = (s',\alpha')$ if mode u adopted. The mobility of user is independent of the remaining data s and the mode u, and thus we have

$$p(e'|e,u) = p\left((s',\alpha'),u|(s,\alpha),u\right) = p\left(\alpha'|\alpha\right)p\left(s'|(s,\alpha),u\right), \quad (5)$$

where

$$p\left(s'|(s,\alpha),u\right) = \begin{cases} 1 & \text{if } s' = [s - R(\alpha,u)(\Delta t - t_b(\alpha,u))]^+ \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

and $[a]^+ = \max\{0,a\}$. The probability for UE to move from location α to location α' is $p\left(\alpha'|\alpha\right)$, which is obtained from the estimation of user's historical mobility pattern.

The selection strategy is defined as modes set $\boldsymbol{\pi} = \{\xi_t(s,\alpha), \forall s \in \mathcal{S}, \forall \alpha \in \mathcal{A}, \forall t \in \mathcal{T}\}$, where $\xi_t : \mathcal{S} \times \mathcal{A} \to \mathcal{U}$ is the mode selection function at state $e = (s,\alpha)$. All $\boldsymbol{\pi}$ compose the feasible set Π. $e_t^{\pi} = (s_t^{\pi}, \alpha_t^{\pi})$ represents the state at time t if strategy $\boldsymbol{\pi}$ is adopted. We try to find the optimal strategy $\boldsymbol{\pi}^*$ to minimize the expected user payment from $t=1$ to $t=T$ and the penalty at $T+1$ [1], [12]:

$$\begin{aligned} &\min_{\pi \in \Pi} E_{e_1}^{\pi}\left[\sum_{t=1}^{T} m_t\left(e_t^{\pi}, \xi_t\left(s_t^{\pi}, \alpha_t^{\pi}\right)\right) + \overline{m}_{T+1}\left(e_{T+1}^{\pi}\right)\right] \\ &\text{s.t. } \xi_t \in \{0,1,2,3\} \\ &\qquad e_1 = (S, \alpha_1), \end{aligned} \quad (7)$$

where $E_{e_1}^{\pi}$ indicates the expectation function in user mobility distribution, and the transmission strategy $\boldsymbol{\pi}$ is subject to an initial state $e_1 = (S,\alpha_1)$, where S is the total data amount to deliver in position α_1 when starting at $t=1$. For simplicity, we assume that the user payment and penalty have equal weights in the cost function. The weight embodies the user's sensitivity to payment or completion.

IV. DELAY-AWARE LTE WLAN AGGREGATION

In this section, based on finite-horizon dynamic programming [26], we propose the Delay-Aware LTE WLAN aggregation algorithm (DLWA) to obtain the optimal transmission strategy $\boldsymbol{\pi}^*$. With the optimal equation in [20], minimal total cost function at state e for $t \in \mathcal{T}$ can be described as

$$w_t(e) = w_t(s,\alpha) = \min_{u \in \mathcal{U}^{(\alpha)}} \{\mu_t(s,\alpha,u)\}, \quad (8)$$

where $\mu_t(s,\alpha,u)$ is the cost function for epoch t

$$\begin{aligned} &\mu_t(s,\alpha,u) \\ &= m_t(s,\alpha,u) + \sum_{\alpha' \in \mathcal{A}} \sum_{s' \in \mathcal{S}} p\left((s',\alpha')|(s,\alpha),u\right) w_{t+1}(s',\alpha') \\ &= \min\{s, R(\alpha,u)(\Delta t - t_b(\alpha,u))\} \cdot \gamma(\alpha,u) \\ &\quad + \sum_{\alpha' \in \mathcal{A}} p\left(\alpha'|\alpha\right) w_{t+1}\left([s - R(\alpha,u)(\Delta t - t_b(\alpha,u))]^+, \alpha'\right). \end{aligned} \quad (9)$$

In the first equation, the total cost from t to $T+1$ is divided into two parts: i) $m_t(s,\alpha,u)$ is the payment for data usage with mode u at t, ii) the second part is the expected follow-up cost in the next slot after choosing mode u. The second equation is by substituting (3)-(6) into the first equation. When exceeding deadline $t = T+1$, the equation contains the penalty function w_{T+1}, which is given as follows:

$$w_{T+1}(e) = \overline{m}_{T+1}(s,\alpha) = \chi(s), \quad \forall s \in \mathcal{S},\ \forall \alpha \in \mathcal{A}. \quad (10)$$

Clearly, if the backhaul latency is intolerable, and aggregation rate is low, i.e., $H(\alpha,3) \leq H(\alpha,1)$, we have

$$\mu_t(s,\alpha,1) \leq \mu_t(s,\alpha,3), \quad \forall s \in \mathcal{S},\ \forall t \in \mathcal{T},\ \alpha \in \mathcal{A}^{(1)}. \quad (11)$$

Proof: See Appendix B. □

It means the cost of LWA mode is larger than that of the scenario where LTE is directly used at that slot due to the poor backhaul condition, and thus the aggregation mode is excluded from the candidates mode set

$$\hat{\mathcal{U}}^{(\alpha)} = \begin{cases} \{0,1\}, & \text{if } \alpha \in \mathcal{A}^{(0)}, \\ \{0,1,2\}, & \text{if } \alpha \in \mathcal{A}^{(1)}. \end{cases} \quad (12)$$

Then, the problem has been simplified to the delayed offloading, which has been discussed in [11]–[13]. From above, it is found that one lower bound of the aggregation rate is identical to $R(\alpha,1)$, i.e., aggregation starts only if the aggregation rate is no less than the LTE average rate. The minimum aggregation gain can be written as:

$$G_{\min} \geq \frac{R(\alpha,1)}{R(\alpha,3)} = \frac{R(\alpha,1)}{R(\alpha,1)+R(\alpha,2)}. \quad (13)$$

In this paper, we mainly focus on the situation when the aggregation gain is above $G_{\min}$ in the following parts. Note that, since the user payment and deadline also need to be considered in mode selection, it is possible that the aggregation may not happen even if the aggregation rate is above the bound.

Theorem 1: the strategy $\boldsymbol{\pi}^* = \{\xi_t{}^*(s,\alpha), \forall s \in \mathcal{S}, \forall \alpha \in \mathcal{A}, \forall t \in \mathcal{T}\}$ *is the optimal solution of problem (7), if*

$$\xi_t^*(s,\alpha) := \arg\min_{u \in \mathcal{U}^{(\alpha)}} \{\mu_t(s,\alpha,u)\}, \tag{14}$$

Proof: See the principle of optimality [26]. □

With the optimality equation in (8) and Theorem 1, we illustrated DLWA algorithm as follows. The first phase of the algorithm is planning the transmission strategy for all slots. Set $\nu = 1$ Mbit as the granularity of the discrete state element s. The optimal transmission strategy $\boldsymbol{\pi}^*$ is obtained by backward induction in solving the objective function in (7) with the optimal optimality in (8) and penalty in (10). Specifically, we first set w_{T+1} as the boundary condition of the algorithm. Then, we obtain the optimal mode for each epoch as $\xi_t^*(s,\alpha)$ by updating them accordingly. The computational complexity of DLWA can be expressed as $\mathcal{O}(SAT/\nu)$ [27]. The pseudo-code of DLWA algorithm is given in Algorithm 1.

Algorithm 1 Delay-Aware LTE WLAN Aggregation Algorithm

```
Plan Phase: Input $S, T, \mathcal{A}$
Set $m_{T+1}(s,\alpha)$, for $\forall t \in \mathcal{T}, \forall \alpha \in \mathcal{A}$ using (10)
Set $t := T$ and begin in recursive backward
while $t > 1$ do
    for $\alpha \in \mathcal{A}$ do
        $s := 0$
        while $s < S$ do
            Calculate $\mu_t(s,\alpha,u), u \in \mathcal{U}^{(\alpha)}$ using (9)
            Set $\xi_t^*(s,\alpha) := \arg\min_{u \in \mathcal{U}^{th}}\{\mu_t(s,\alpha,u)\}$
            $w_t(s,\alpha) = \mu_t(s,\alpha,\xi_t^*(s,\alpha))$
            $s := s + \nu$
        end while
    end for
    $t := t + 1$
end while
Output the optimal strategy $\boldsymbol{\pi}^*$.
```

V. THRESHOLD-BASED DELAY-AWARE LTE WLAN AGGREGATION ALGORITHM

In this section, we reveal that a threshold structure of LTE usage exists in the dimension of the remaining data s and time t, and further develop the low-computational complexity threshold-based DLWA (TB-DLWA) based on this structure. The following assumptions are made to derive the threshold structure.

Assumption: (a) The penalty $\chi(s)$ is convex, and non-decreasing in time slot t and remaining sequence s; (b) The LTE usage price per unit is location independent, i.e., $\gamma(\alpha,1) = \gamma(\alpha',1), \forall \alpha,\alpha' \in \mathcal{A}^{(1)}, \alpha \neq \alpha'$. (c) WiFi is free as incentives, i.e., $\gamma(\alpha,2) = 0, \forall \alpha \in \mathcal{A}^{(1)}$; (d) The WiFi data rate $R(\alpha,2)$ is location-independent. Thus, the throughput of WiFi network achieved in one slot is $H_2 = H(\alpha,2) = R(\alpha,2)\Delta t, \ \forall \alpha \in \mathcal{A}^{(1)}$; (e) Since only consider the large scale fading, the LTE network provides the same rate for users with the similar distance to the eNB, i.e., user in square index belonging to the same set $I_1 = \{1,4,13,16\}$, $I_2 = \{6,7,10,11\}$ and $I_3 = \{2,3,5,8,9,12,14,15\}$ have the same LTE throughput. Then, the throughput of LTE network can be quantized as several values in the set $H_1 \in \{H_{1i}|H_{1i} = R(\alpha,1)\Delta t, \alpha \in I_i, i = 1,2,3\}$; (f) The latency for non-idea backhaul is location-independent, i.e., $t_b = t_b(\alpha,u)$. The throughput for LWA in one slot could be rewritten as $H_3 = H(\alpha,3) = [R(\alpha,1) + R(\alpha,2)] \cdot (\Delta t - t_b) = (H_1 + H_2) \cdot G$, where $G(\alpha,u) = \frac{\Delta t - t_b}{\Delta t}$, which reflects the link status between LTE and WiFi. Particularly, we have $H_3 = H_1 + H_2$ for a perfect backhaul.

A. MONOTONE PROPERTIES OF THE OPTIMAL EQUATION

With these assumptions, it is found that users only pay for the LTE usage. Let l be the indicator of LTE usage.

$$l = \begin{cases} 1 & u = 1 \text{ for } \alpha \in \mathcal{A}^{(0)} \text{ or } u = 3 \text{ for } \alpha \in \mathcal{A}^{(1)}, \\ 0 & u = 0 \text{ for } \alpha \in \mathcal{A}^{(0)} \text{ or } u = 2 \text{ for } \alpha \in \mathcal{A}^{(1)}. \end{cases} \tag{15}$$

Namely, the $l = 1$ holds when the modes $u = 1$ for LTE direct usage, and $u = 3$ for the LTE usage in aggregation. Then, the payment function in (3) could be written as

$$m_t(e,u) = m_t(s,\alpha,u) = \delta(l=1)\,\gamma_u = \begin{cases} \gamma_u & \text{if } l = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

where $\forall \alpha \in \mathcal{A}^{(\alpha)}$, and $\delta(\cdot)$ is the indicator function. $\gamma_1 = \gamma(\alpha,1)H_1$, while the payment for LWA only relies on the use of LTE $\gamma_3 = \gamma(\alpha,1)\,(H_3 - H_2G)$. As a result, payment function in (9) can be rewritten as

$$\mu_t(s,\alpha,u) = \delta(l=1)\gamma_u + \sum_{\alpha' \in \mathcal{A}} p\left(\alpha'|\alpha\right) w_{t+1}([s - H(\alpha,u)]^+, \alpha'), \tag{17}$$

It is shown that the tradeoff between user payment and QoS depends on the LTE usage. To be more specific, the LTE usage increases the completion probability and user payment at the same time. Thus, the LTE usage should be carefully scheduled, which motivates us to deduce the LTE usage structure. Inspired by [12] and [25], the optimal strategy has a threshold structure with the convex penalty function as shown in Assumption (a). In the following parts, we give the threshold structure in LTE usage likewise.

Proposition 1: the monotone properties of optimal equation (minimum cost function) $w_t(s,\alpha)$*:*

(a) $w_t(s,\alpha)$ *is a non-decreasing with s,* $\forall \alpha \in \mathcal{A}, \forall t \in \mathcal{T}$.

(b) $w_t(s,\alpha)$ *is a non-decreasing with t,* $\forall \alpha \in \mathcal{A}, \forall s \in \mathcal{S}$.

Proof: See Appendix A. □

Intuitively,a larger amount of data s usually means a higher expected payment, while larger t indicates less time for transmission and LTE is more likely to be used.

B. THRESHOLD STRUCTURE OF LTE USAGE

In this subsection, we derive the threshold structure for LTE usage. Firstly, we characterize the optimal mode at location $\alpha \in \mathcal{A}^{(1)}$ with WiFi.

Proposition 2: For $\alpha \in \mathcal{A}^{(1)}$ $\forall s \in \mathcal{S}$, $\forall t \in \mathcal{T}$, we have:

$$\mu_t(s,\alpha,0) \geq \mu_t(s,\alpha,2),\ \ \mu_t(s,\alpha,1) \geq \mu_t(s,\alpha,3), \quad (18)$$

which is proved in Appendix B. From *Proposition 2*, we find the WiFi network has higher priority to be used as it could reduce the payment, and the mode set $\mathcal{U}^{(\alpha)}$ in (2) is simplified as:

$$\tilde{\mathcal{U}}^{(\alpha)} = \begin{cases} \{0,1\}, & \text{if } \alpha \in \mathcal{A}^{(0)}, \\ \{2,3\}, & \text{if } \alpha \in \mathcal{A}^{(1)}. \end{cases} \quad (19)$$

Thus, the minimum cost function is

$$w_t(s,\alpha) = \min_{u\in\mathcal{U}^{(\alpha)}} \{\mu_t(s,\alpha,u)\} = \min_{u\in\tilde{\mathcal{U}}^{(\alpha)}} \{\mu_t(s,\alpha,u)\}. \quad (20)$$

It can be observed from (19) that only two candidate modes in both subsets after simplification, which reduces the searching complexity in (20). It is found that the difference of the two candidates in the same subset is whether to use the LTE, for example, WLAN offloading ($u = 2$) differs with LWA ($u = 3$) in whether to aggregate with LTE resource. As results, this observation motivates to deduce the threshold in LTE usage.

1) THRESHOLD STRUCTURE OF LTE USAGE IN REMAINING DATA SIZE

Firstly, we derive the subadditivity properties [20] of cost function $\mu_t(s,\alpha,u)$ in Appendix C. The subadditivity properties is:

Definition 1: For given $\alpha \in \mathcal{A}$, $\mu_t(s,\alpha,u)$ is subadditive on $\mathcal{S} \times \tilde{\mathcal{U}}^{(\alpha)}$ if for $\forall s^+, s^- \in \mathcal{S}$ and $\forall u^+, u^- \in \mathcal{U}$, where $s^+ \geq s^-$ and $u^+ \geq u^-$, we have

$$\mu_t(s^+,\alpha,u^+) + \mu_t(s^-,\alpha,u^-) \leq \mu_t(s^+,\alpha,u^-) + \mu_t(s^-,\alpha,u^+). \quad (21)$$

The threshold structure of the LTE usage is as follows:

Theorem 2: the optimal policy $\boldsymbol{\pi}^* = \{\xi_t{}^*(s,\alpha), \forall s \in \mathcal{S}, \forall \alpha \in \mathcal{A}, \forall t \in \mathcal{T}\}$ has a threshold structure of LTE usage in s. For $\forall t \in \mathcal{T}$, the LTE usage indication function l has

$$l(s,t) = \begin{cases} 1 & \text{if } s \geq s^*(\alpha,t), \\ 0 & \text{otherwise}, \end{cases} \quad (22)$$

and the transmission mode function has:

$$\xi_t{}^*(s,\alpha) = \begin{cases} 1 & \text{(LTE)}, \ \text{if } s \geq s^*(\alpha,t), \\ 0 & \text{(idle)}, \text{otherwise}, \end{cases} \alpha \in \mathcal{A}^{(0)} \quad (23)$$

$$\xi_t{}^*(s,\alpha) = \begin{cases} 3 & \text{(LWA)}, \quad \text{if } s \geq s^*(\alpha,t), \\ 2 & \text{(offloading)}, \text{otherwise}, \end{cases} \alpha \in \mathcal{A}^{(1)} \quad (24)$$

where $s^*(\alpha,t)$ is the threshold in dimension s, which is dependent of location and time.

Proof: See Appendix C and D. □

2) THRESHOLD STRUCTURE OF LTE USAGE IN REMAINING TIME

Theorem 3: the optimal policy $\boldsymbol{\pi}^* = \{\xi_t{}^*(s,\alpha), \forall s \in \mathcal{S}, \forall \alpha \in \mathcal{A}, \forall t \in \mathcal{T}\}$ has a threshold structure of LTE usage in t. For $\forall s \in \mathcal{S}$, the LTE usage indication function l has

$$l(s,t) = \begin{cases} 1 & \text{if } t \geq t^*(\alpha,s), \\ 0 & \text{otherwise}, \end{cases} \quad (25)$$

and the transmission mode function has:

$$\xi_t{}^*(s,\alpha) = \begin{cases} 1 & \text{(LTE)}, \quad \text{if } t \geq t^*(\alpha,s), \\ 0 & \text{(idle)}, \quad \text{otherwise}, \end{cases} \alpha \in \mathcal{A}^{(0)} \quad (26)$$

$$\xi_t{}^*(s,\alpha) = \begin{cases} 3 & \text{(LWA)}, \quad \text{if } t \geq t^*(\alpha,s), \\ 2 & \text{(offloading)}, \text{otherwise}, \end{cases} \alpha \in \mathcal{A}^{(1)} \quad (27)$$

where $t^*(\alpha,s)$ is the threshold in dimension t, which is dependent of location and data size.

Proof: See Appendix E and F. □

Theorem 2 and 3 reveal the threshold structure of LTE usage exists in dimensions s and t. With this, the optimal strategy could be concluded as: it is optimal to use the LTE resource to avoid penalty

a) when remaining data s exceeds the LTE usage threshold $s^*(s,\alpha)$, i.e., the remaining data is beyond the capacity of WiFi network within given time;

b) when t oversteps the LTE usage threshold $t^*(\alpha,s)$, i.e., the time left is not enough for transmission without using LTE resource.

Theorem 2 and 3 is noteworthy since simplified algorithm could be developed by comparing with the threshold structure and avoiding complicated backward induction in dynamic programming [25].

3) PROPERTIES IN THRESHOLD STRUCTURE

Further, we give the properties in threshold structure like [12] and [25] to simplify the threshold generation process.

Theorem 4: For $\forall t \in \mathcal{T}$, the threshold in s, $s^*(\alpha,t)$, is the non-increasing function in t.

$$s^*(\alpha,t-1) \geq s^*(\alpha,t),\ \forall \alpha \in \mathcal{A}. \quad (28)$$

For $\forall s \in \mathcal{S}$, the threshold in t, $t^*(s,\alpha)$, is non-increasing function in s

$$t^*(s,\alpha) \geq t^*(s+\nu,\alpha),\ \forall \alpha \in \mathcal{A}. \quad (29)$$

Proof: See Appendix G. □

C. THRESHOLD-BASED DELAY-AWARE LTE WLAN AGGREGATION ALGORITHM

In this subsection, we develop Threshold-based Delay-Aware LTE WLAN Aggregation (TB-DLWA) algorithm, which is illustrated in Algorithm 2.

The difference in TB-DLWA is the optimal solution set $\boldsymbol{\pi}^*$ is obtained by comparing the state with the LTE usage threshold set $\{s^*(\alpha,t), \forall \alpha \in \mathcal{A}, \forall t \in \mathcal{T}\}$ rather than backward induction. The threshold process generates from

Algorithm 2 Threshold-Based Delay-Aware LTE WLAN Aggregation Algorithm

```
Plan Phase: Input $S$, $T$, $\mathcal{A}$
Set $m_{T+1}(s, \alpha)$, for $\forall t \in \mathcal{T}$, $\forall \alpha \in \mathcal{A}$ using (10)
Set $t := T$
while $t > 1$ do
    for $\alpha \in \mathcal{A}$ do
        Call Threshold Process
    end for
    $t := t - 1$
end while
Output the threshold set $\{s^*(\alpha, t),\ \forall \alpha \in \mathcal{A}, \forall t \in \mathcal{T}\}$
Transmission Phase:
Set $t := 1$ and $s := S$
while $t \leq T$ do and $s > 0$
    Get area index $\alpha$
    if $s \geq s^*(\alpha, t)$ then
        Set $u := 1$ for $\alpha \in \mathcal{A}^{(0)}$, or $u := 3$ for $\alpha \in \mathcal{A}^{(1)}$
    else
        Set $u := 0$ for $\alpha \in \mathcal{A}^{(0)}$, or $u := 2$ for $\alpha \in \mathcal{A}^{(1)}$
    end if
    $t := t + 1$
end while
```

Algorithm 3 Threshold Process

```
function Threshold Process
    Set $\mathcal{U}_0^{(\alpha)} = \{0, 1\}$ for $\alpha \in \mathcal{A}^{(0)}$; and $\mathcal{U}_1^{(\alpha)} = \{2, 3\}$
    for $\alpha \in \mathcal{A}^{(0)}$;
    Initialize $s := 0$ and $flag := 0$
    while $s \leq S$ do
        if $s \geq s^*(\alpha, t + 1)$ and $flag := 0$ then
            Set $flag := 1, \mathcal{U}_0^{th} = \{1\}$ for $\alpha \in \mathcal{A}^{(0)}$, or
            $\mathcal{U}_1^{th} = \{3\}$ for $\alpha \in \mathcal{A}^{(1)}$
        end if
        Calculate $\mu_t(s, \alpha, u)$, $u \in \mathcal{U}^{th}$ using (9)
        Set $xi_t^*(s, \alpha) := \arg\min_{u \in \mathcal{U}^{th}}\{\mu_t(s, \alpha, u)\}$
        $w_t(s, \alpha) = \mu_t(s, \alpha, \xi_t^*(s, \alpha))$
        $l = \delta\left[\xi_t^*(s, \alpha)\right]$
        if $l = 1$ then
            Set $\mathcal{U}_0^{th} = \{1\}$ for $\alpha \in \mathcal{A}^{(0)}$, or $\mathcal{U}_1^{th} = \{3\}$
            for $\alpha \in \mathcal{A}^{(1)}$
            Set $flag := 1$ and $s^*(\alpha, t) := s$
        end if
        $s := s + \nu$
    end while
end function
```

Algorithm 3 at the planning phase, and it could be further simplified. In particular, since the backtracking process in threshold generation, the search space of $s^*(\alpha, t-1)$ can be reduced due to the monotonicity of threshold in Theorem 4, if $s^*(\alpha, t)$ is obtained. With the threshold structure, we obtain optimal mode set $\boldsymbol{\pi}^*$ by only considering the unique mode instead of two candidates in the subset. To be more specific:

(i) When $s < s^*(\alpha, t)$, it is not wise to use LTE resources ($l^* = 0$). Therefore, we only choose $\mathcal{U}_0^{th} = \{0\}$ for $\alpha \in \mathcal{A}^{(0)}$ or $\mathcal{U}_1^{th} = \{2\}$ for $\alpha \in \mathcal{A}^{(1)}$, rather than consider both candidates in $\tilde{\mathcal{U}}^{(\alpha)}$.

(ii) When the remaining data is larger than the threshold, i.e., $s > s^*(\alpha, t)$, LTE resource should be used ($l^* = 1$), to be specific, $\mathcal{U}_0^{th} = \{1\}$ for $\alpha \in \mathcal{A}^{(0)}$ or aggregation $\mathcal{U}_1^{th} = \{3\}$ for $\alpha \in \mathcal{A}^{(1)}$.

Both cases in (i) and (ii) show the threshold-based subset $\mathcal{U}^{th}$ with a single element. In the transmission phase, actions are made based on the optimal selection set $\boldsymbol{\pi}^*$. The complexity of TB-DLWA drops from $\mathcal{O}\,(SAT/\nu)$ in DLWA to approximately $\mathcal{O}\,(A \cdot \max\{S/\nu, T\})$ [27].

VI. PERFORMANCE EVALUATION

The proposed DLWA and TB-DLWA algorithms are evaluated in this section. As shown in Fig.1, WLAN APs are randomly located in the 16 squares with density ρ_w, and no overlap and interference with each other. The simulation results are obtained by randomized AP locations averaging over 1000 simulating runs. The parameter settings are: $N_0 = -80$dBm, $p\left(\alpha'|\alpha\right) = 0.6$, $\alpha_0 = 2$, the granularity $\nu = 1$ Mbyte, the square in Fig. 1 is with a length of 100m. The length of time slot Δt is one second. Price for LTE data usage is \$10/GByte. The convex penalty function for not finishing is defined as $\chi(s) = 10s^2, \forall s \in \mathcal{S}$ [12]. Other assumptions and detailed parameters are summarized in Table 1.

A. PERFORMANCE COMPARISON WITH DIFFERENT SCHEMES WITH IDEAL BACKHAUL

In this section, we consider an ideal backhaul ($G = 1$) for LWA, and compare the proposed algorithms with other LTE WLAN interworking schemes under different kinds of deadlines. The results are obtained under the mean WiFi rate $R_w = 40$ Mbps and AP density $\rho_w = 0.6$. Comparisons are made with the following benchmarks (1) WLAN Preferred (WP), (2) 3GPP Release 12 WLAN interworking solution (Rel-12 interworking), (3) delayed WiFi offloading, (4) Always LWA Aggregation. WLAN Preferred scheme enables the UE connects to the WLAN networks whenever WiFi available. For Rel-12 interworking, the UE associates to the WLAN only when the data rate of LTE below a certain threshold. The optimal value for this threshold is empirically found in [18]. WP and Rel-12 are opportunistic offloading schemes. We also compare the proposed algorithms with the delayed WiFi offloading [11], [12]. At last, we consider the Always Aggregation scheme [17]–[19], which performs the LWA whenever possible without consideration of the deadline. We divide the deadline into several categories, including tight, moderate and loose deadlines. The tight deadline means the file transmission is difficult to complete even with full LTE usage within the transmission deadline. We then focus on the completion probability in this case.

TABLE 1. Simulation parameters.

LTE	
Parameters	Description
Topology	3 sector per macrocell, 5 small cell/ secor, 7 cell wrap-around,Small cell LTE uses same carrier as macro-cell, No ICIC
Carrier Frequency	2 GHz
Channel / UE Speed	[IMT] UMa Macro, UMi Pico UE speed= 10km/hr
LTE mode	Downlink FDD; 20 MHz for DL
Transmission power	25 dBm
UE channel estimation	Ideal
WLAN	
Topology	IEEE 802.11n based APs, no overlap random distribution in deployment density
WiFi Frequency	2.4 GHz band, 3 frequency bands,
Channelization	20 MHz channels; least power based channel selection
AP Transmit power	20 dBm outdoor
WiFi mode	Downlink only
TX-OP	1ms

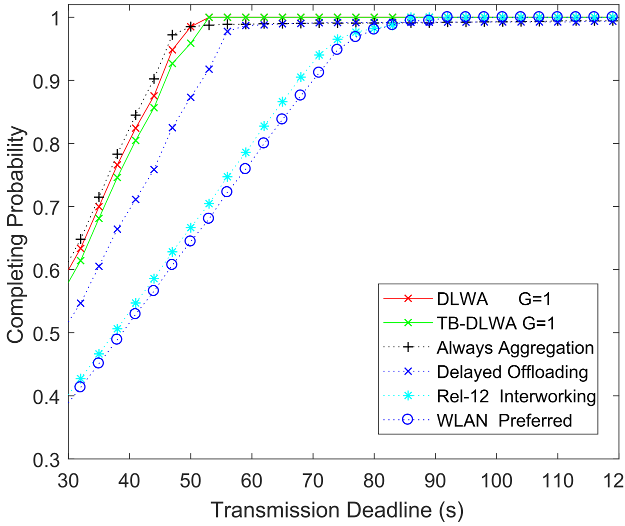


FIGURE 2. Completion probability versus short transmission deadline (S = 900 MBytes, $T_{\max}$ = 120s).

Meanwhile, the loose deadline indicates the transmission can be finished only by using low-rate WiFi networks. Since the completion is guaranteed in loose deadline, the main concern should be the user payment. Between two particular cases, where the completion can be achieved by partially using LTE resources, we classify it into moderate deadline situation.

First, Fig. 2 illustrates the completion probability of each scheme in the tight deadline, e.g. large file $S = 900$ MBytes needs to be delivered within a short time ($T < 120$s). With the increasing of T, the completion probability of all schemes increase. Moreover, it can be observed that the proposed DLWA algorithms and Always Aggregation scheme achieve the highest completion probability, followed by the TB-DLWA scheme. The reason is, with LWA, resources in both LTE and WiFi network could be used for transmission simultaneously, and thus boosts the data rate and guarantee higher completion. On the other hand, we observe that opportunistic WiFi offloading schemes, such as WLAN Prefer or Re-12 interworking, fall to accomplish the transmission with below 50% completion probability when T = 40s. It is because these two schemes always offload the data to the WiFi networks without consideration of QoS requirement. Delayed offloading scheme outperforms to the opportunistic offloading schemes in completion probability. Since data rate of LTE is usually larger than that of WiFi, delayed offloading inclines to LTE mode to avoid the unaccomplished penalty despite of higher payment.

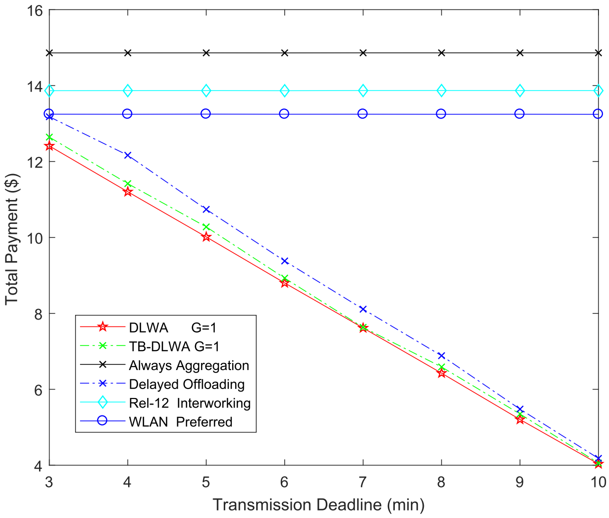


FIGURE 3. Total user payment versus short transmission deadline (S = 900 MBytes, $T_{\max}$ = 10 mins).

Next, we evaluate the performance in moderate deadline as shown in Fig. 3. As mentioned above, when the deadline is sufficient to complete the transmission, the user payment is concerned in this situation. When the deadline is longer, DLWA and delayed offloading scheme use the longer delay tolerance to wait for WiFi networks, and reduce their LTE data usage, and thus the payment decreases with the transmission deadline as shown in Fig. 3. Moreover, for DLWA, when the deadline is large enough (T >10 mins), the optimal mode is WiFi offloading, and thus DLWA achieves the similar results as delayed offloading. Without consideration of delay, the schemes like Always Aggregation and opportunistic WiFi offloading, the user payment for these schemes are independent of T. Since the greedy method adopted by Always Aggregation, it results in high user payment. From Fig. 2 and Fig. 3, it can be concluded that the proposed algorithms could

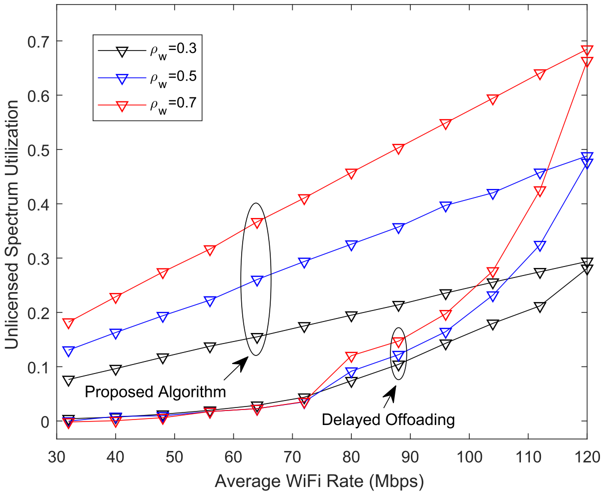


FIGURE 4. **Unlicensed spectrum utilization versus average WiFi rate for moderate deadline (S = 750 MBytes, T = 50s).**

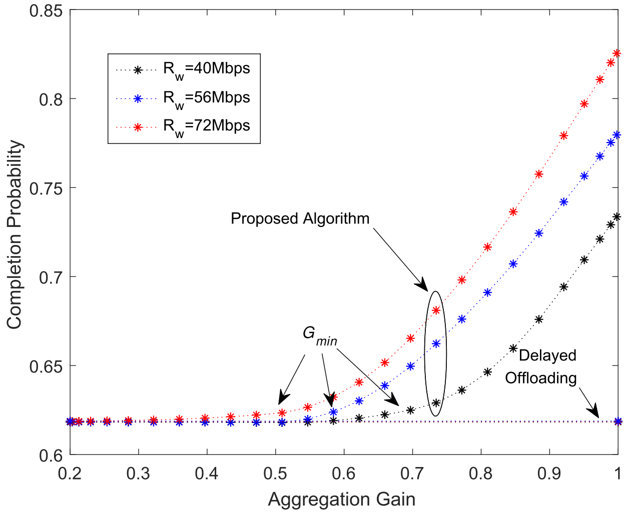


FIGURE 5. **Completion probability versus aggregation gain G under different WiFi average rates R_w for WiFi density ρ_w = 0.6 in tight deadline (S = 750 MBytes, T = 30s).**

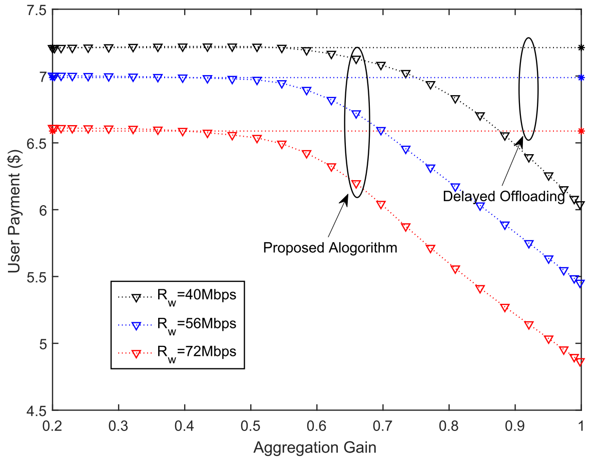


FIGURE 6. **User payment versus aggregation gain G under different WiFi average rates R_w for WiFi density ρ_w = 0.6 in moderate deadline (S = 750MBytes, T = 50s).**

achieve a better trade-off in user payment and QoS compared with the current integrated cellular and WiFi schemes.

In Fig. 4, we further compare the proposed algorithm and delayed offloading scheme in term of unlicensed spectrum utilization, which is defined as the ratio of the amount of data delivered by WiFi network to the total amount. As we can see, when $R_w < 70$ Mbps, the delayed offloading lead to low efficiency in unlicensed spectrum utilization. It chooses LTE mode for high completion rather than use WiFi networks, which also accounts for the better performance than other offloading schemes in Fig. 1. However, this nevertheless leads to higher user payment and cannot make the best use of unlicensed spectrum. On the contrary, it can be observed that the DLWA has a sustained increase in unlicensed utilization with the larger R_w and ρ_w. With the dual connection in DLWA, all the benefits bought by larger WiFi density ρ_w or rate R_w could be used.

B. IMPACTS OF NON-IDEAL BACKHAUL ON AGGREGATION

In this section, we investigate the impacts of the non-ideal backhaul on aggregation decision and mode selection by comparing with the delayed offloading scheme [11], [12]. Notice that, without the aggregation mode, the performance of the delayed offloading scheme is independent of the aggregation gain G.

In Fig. 5, we plot the completion probability varies with aggregation gain G under different WiFi average rates R_w in tight deadline (e.g. $S = 750$ Mbytes, $T = 30$s). It can be observed that, for given R_w, the completion probability increases with the aggregation gain. Intuitively, when the aggregation starts, better aggregation gain contributes to the throughput, and thus it is more likely to finish the transmission. However, in the low aggregation gain region ($G < 0.5$), the performance of DLWA is the almost same with delayed offloading. Recall that it is indicated in (12) that the LWA mode $u = 3$ would be excluded in the candidate set $\mathcal{U}^{(\alpha)}$ when the aggregation gain $G < G_{\min}$, and the problem be simplified as the delayed offloading. From the separation point of two curves, it can be observed that the completion probability increases when $G \geq G_{\min}$. Moreover, $G_{\min}$ decreases with the R_w, which verifies that minimum aggregation gain $G_{\min}$ is determined by the ratio of average LTE and WiFi data rate in (13), e.g., for given average LTE data rate, the lower bound of $G_{\min}$ decreases with average WiFi data rate R_w.

Fig. 6 illustrates the user payment varies with aggregation gain under different WiFi average rates R_w in moderate deadline ($S = 750$ Mbytes, $T = 50$s). It can be observed from Fig. 6 that the user payment decreases with the aggregation gain. It is because, when $G \geq G_{\min}$, DLWA could dynamically choose LWA mode or offloading mode whenever possible, and thus reduce the LTE usage.

To further illustrate the impacts of imperfect backhaul, we plot aggregation ratio varies with aggregation gain G

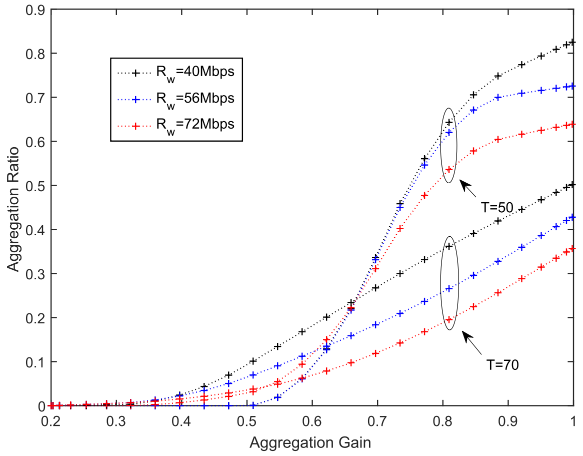


FIGURE 7. Aggregation ratio versus aggregation gain G under different WiFi average rates R_W for $\rho_W = 0.6$, $S = 750$MBytes, $T = 50, 70$ s.

under different WiFi average rates R_w when $S = 750$ Mbytes, $T = 50$s, $T = 70$s in Fig. 7. It can be observed that, for given T and R_w, LWA are more favorable. With better aggregation gain G, LWA could use the network resource effectively, and is more likely to be the optimal mode. From the different deadlines, we observe that the aggregation ratios for $T = 70$s are lower than $T = 50$s. It is because the DLWA could dynamically determine the aggregation, and prefers WiFi offloading mode to reduce LTE usage when given larger deadline T, which turns the aggregation ratio down. In summary, the proposed DLWA algorithm could reach the balance between user payment and QoS requirement, and delayed offloading scheme could be deemed as the low bound for the proposed algorithm.

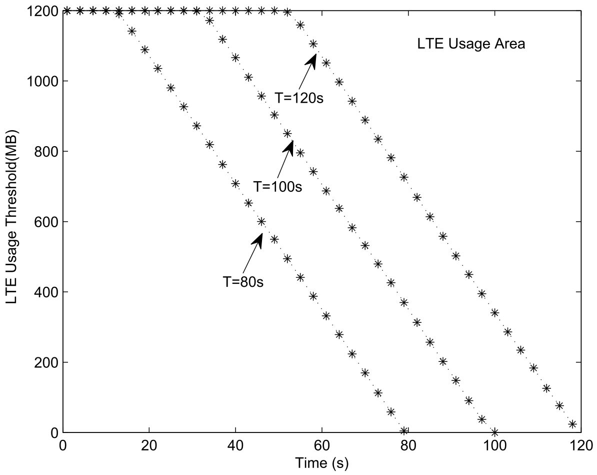


FIGURE 8. Threshold structure of LTE usage versus different deadline constraint.

C. DEMONSTRATION OF THRESHOLD STRUCTURE OF LTE USAGE

In this subsection, we demonstrate the threshold structure of LTE usage. First, Fig. 8 illustrates the LTE usage threshold for delivering $S = 1.2$ GByte data under different T. The LTE usage threshold in s, $s^*(\alpha, t)$, is depicted by the curves in Fig. 8. The area above the curves, i.e., $s > s^*(\alpha, t)$, is defined as LTE usage area, which means it is optimal to use the LTE network if the state locates in this are, i.e., $l^* = 1$. Moreover, it is observed that, as deduced in Theorem 4, $s^*(\alpha, t)$ is non-increasing in t.

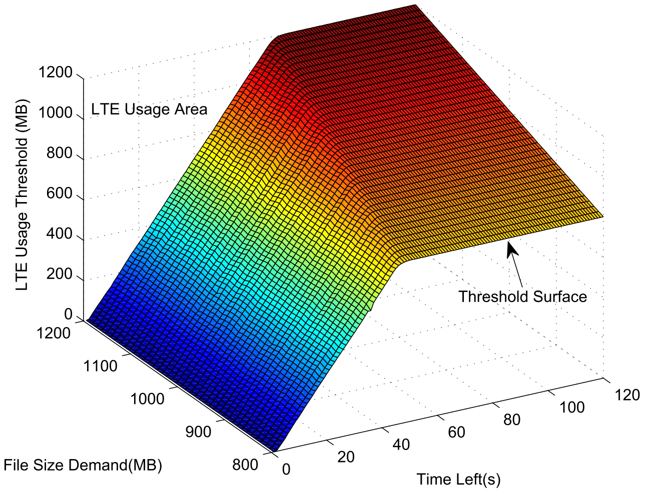


FIGURE 9. Threshold structure for LTE usage versus different transmission sizes.

Furthermore, the threshold of LTE usage for different sizes of data transmission under the deadline $T = 120$ s is illustrated in Fig. 9. When the remaining file $s \geq s^*(\alpha, t)$ or time $t \geq t^*(s, \alpha)$, the LTE network should be used (e.g. $l^* = 1$). These states locate in the space above the threshold surface in Fig. 9. Thus, we verify that the LTE usage depends on the time left and remaining data size. Moreover, from the axis of LTE usage threshold, it is found that the LTE usage threshold $t^*(s, \alpha)$ is the non-increasing in s as Theorem 4.

VII. CONCLUSION

In this paper, we study the integrated LTE and WiFi network access problem, which aims to minimize data usage payment with consideration of the deadline. It is shown that, by jointly considering LWA with WiFi offloading, the proposed algorithm could achieve a better tradeoff in user payment and QoS requirement compared with the current cellular and WiFi interworking schemes. When the deadline is tight, the proposed algorithm could increase the data rate by using carrier aggregation in LWA, and thus guarantee the QoS requirement. While the deadline is loose, it could reduce the LTE usage by WiFi offloading, and thus lower the user payment. Moreover, with backhaul, LWA is chosen as optimal mode only when the aggregation gain is larger than the minimum value, which is determined by the average LTE and WiFi rate. It should be noted that the proposed algorithm only considers transmission demand for the single user. For future work, we will explore the transmission strategy for multi-user scenario, and take the congestion problem into the aggregation decision.

APPENDIX A
PROOF OF PROPOSITION 1

Proof: (a) We prove it by induction. First, from $T+1$, $w_{T+1}(s,\alpha)=\chi(s)$ is non-decreasing in s. Assume $w_{t+1}(s,\alpha)$ is a non-decreasing in s. From (17), since $p(\alpha'|\alpha)\geq 0$ and function γ_u is independent of s, so $\mu_t(s,\alpha,u)$ is a non-decreasing function in s. Moreover, from (8), $w_t(s,\alpha)$ is the minimum function of $\mu_t(s,\alpha,u)$, and the minimum property will not change its monotonicity of s. Thus, $w_t(s,\alpha)$ is a non-decreasing function in s.

(b) We prove it by induction. From $T+1$, we have

$$\begin{aligned} w_T(s,\alpha) &= \min_{u\in\mathcal{U}^{(\alpha)}}\{\mu_T(s,\alpha,u)\}\leq \mu_T(s,\alpha,0)\\ &= \sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right) w_{T+1}(s,\alpha')\\ &= \chi(s)=w_{T+1}(s,\alpha), \end{aligned}\tag{30}$$

where the first equation is by the definition of $w_t(s,\alpha)$ in (8).

Similar to (a), suppose for $\forall s\in\mathcal{S}$, $\forall\alpha\in\mathcal{A}$, $w_{t+1}(s,\alpha)$ is non-decreasing in t. From (14), since $p(\alpha'|\alpha)\geq 0$, $\forall\alpha,\alpha'\in\mathcal{A}$, and the function and γ_u is independent of t. Thus, $\mu_t(s,\alpha,u)$ is a non-decreasing function in t. From (8), the minimum of $\mu_t(s,\alpha,u)$ is $w_t(s,\alpha)$, and the minimum property will not change its monotonicity about t. Thus, $w_t(s,\alpha)$ is also the non-decreasing function in t. □

APPENDIX B
PROOF OF PROPOSITION 2

Proof: (1) For $s\in\mathcal{S}$, $\alpha\in\mathcal{A}$, we prove $\mu_t(s,\alpha,1)\leq\mu_t(s,\alpha,3)$ when $H(\alpha,1)\geq H(\alpha,3)$. We have

$$\begin{aligned} &\mu_t(s,\alpha,3)\\ &\quad=\gamma_3+\sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right) w_{t+1}([s-H(\alpha,3)]^+,\alpha)\\ &\quad\geq\gamma_1+\sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right) w_{t+1}([s-H(\alpha,1)]^+,\alpha')\\ &\quad=\mu_t(s,\alpha,1). \end{aligned}\tag{31}$$

where the inequality is based on the property that the $w_{t+1}(s,\alpha)$ is a non-decreasing function in s and fact the cost for multiple slots is larger than the user payment at one slot.

(2) Then, prove $\mu_t(s,\alpha,1)\geq\mu_t(s,\alpha,3)$, and $\mu_t(s,\alpha,0)\geq\mu_t(s,\alpha,2)$ when $H(\alpha,1)\leq H(\alpha,3)$.

$$\begin{aligned} \mu_t(s,\alpha,1) &= \gamma_1+\sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right) w_{t+1}([s-H(\alpha,1)]^+,\alpha)\\ &\geq\gamma_1+\sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right) w_{t+1}\\ &\quad\times\left([s-(H(\alpha,1)+H(\alpha,2)\,G)]^+,\alpha'\right)\\ &\geq\gamma_3+\sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right) w_{t+1}([s-H(\alpha,3)]^+,\alpha')\\ &=\mu_t(s,\alpha,3), \end{aligned}\tag{32}$$

where the former equalities are due to (14) and the inequality is by the non-decreasing property of $w_{t+1}(s,\alpha)$ in s. Similarly,

$$\begin{aligned} \mu_t(s,\alpha,0) &= \sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right) w_{t+1}(s,\alpha)\\ &\geq\sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right) w_{t+1}([s-H(\alpha,2)]^+,\alpha')\\ &=\mu_t(s,\alpha,2), \end{aligned}\tag{33}$$

which completes the proof of (18). □

APPENDIX C
SUBADDITIVITY OF $\mu_t(s,\alpha,u)$

Proof: The proof of the threshold structure in dimension s in Theorem 2 is based on the results in subadditivity of $\mu_t(s,\alpha,u)$, which is illustrated in *Lemma 1* and *Lemma 2*. For $\forall\alpha\in\mathcal{A}$, let $\tilde{\mathcal{U}}^{(\alpha)}=\{0,1\}$ for $\alpha\in\mathcal{A}^{(0)}$, and $\tilde{\mathcal{U}}^{(\alpha)}=\{2,3\}$ for $\alpha\in\mathcal{A}^{(1)}$. With only two possible modes in $\tilde{\mathcal{U}}^{(\alpha)}$. Thus, (14) can be rewritten as (34) in the bottom of this page.

Lemma 1: With $0\leq H_0\leq H_1$, $0\leq H_2\leq H_3$, $\forall 0<\varepsilon\leq\nu$, *if* $\chi(s)$ *is a convex and non-decreasing function in s, then*

(a) For $\alpha\in\mathcal{A}^{(0)}$, $\tilde{\mathcal{U}}^{(\alpha)}=\{0,1\}$

$$\begin{aligned} &w_t([s-H_0]^+,\alpha)-w_t([s-H_1]^+,\alpha)\\ &\quad\geq w_t([s-\varepsilon-H_0]^+,\alpha)-w_t([s-\varepsilon-H_1]^+,\alpha), \end{aligned}\tag{35}$$

(b) For $\alpha\in\mathcal{A}^{(1)}$, $\tilde{\mathcal{U}}^{(\alpha)}=\{2,3\}$

$$\begin{aligned} &w_t([s-H_2]^+,\alpha)-w_t([s-H_3]^+,\alpha)\\ &\quad\geq w_t([s-\varepsilon-H_2]^+,\alpha)-w_t([s-\varepsilon-H_3]^+,\alpha). \end{aligned}\tag{36}$$

Proof: Since intercommunity exists in (a) and (b), we mainly give the proof of (a), while (b) could be obtained in similar approaches. Following [12] and [25], we prove it by induction. Since $\chi(s)$ is a non-decreasing and convex about s, for $\forall s\in\mathcal{S}$,

$$\begin{aligned} &\chi([s-H_0]^+)-\chi([s-H_1]^+)\\ &\qquad\geq\chi([s-\varepsilon-H_0]^+)-\chi([s-\varepsilon-H_1]^+). \end{aligned}\tag{37}$$

We induce from $t=T+1$, and then we have

$$\begin{aligned} &w_{T+1}([s-H_0]^+,\alpha)-w_{T+1}([s-H_1]^+,\alpha)\\ &\quad=\chi([s-H_0]^+)-\chi([s-H_1]^+) \end{aligned}$$

$$\begin{aligned} \mu_t(s,\alpha,u) &= \delta(l=1)\gamma_u\\ &\quad+\sum_{\alpha'\in\mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right)\left[\delta(l=1)w_{t+1}([s-H_1)]^+,\alpha')+(1-\delta(l=1))\,w_{t+1}([s-H_0)]^+,\alpha')\right]\\ &\quad+\sum_{\alpha'\in\mathcal{A}^{(1)}} p\left(\alpha'|\alpha\right)\left[\delta(l=1)w_{t+1}([s-H_3)]^+,\alpha')+(1-\delta(l=1))\,w_{t+1}([s-H_2)]^+,\alpha')\right] \end{aligned}\tag{34}$$

$$\begin{aligned} &\geq \chi([s-\varepsilon-H_0]^+) - \chi([s-\varepsilon-H_1]^+) \\ &= w_{T+1}([s-\varepsilon-H_0]^+,\alpha) - w_{T+1}([s-\varepsilon-H_1]^+,\alpha), \end{aligned} \tag{38}$$

where the equalities are by (10), and the inequality is by (34). Assume for a given $t \in \mathcal{T}$, we have:

$$\begin{aligned} &w_{t+1}([s-H_0]^+,\alpha) - w_{t+1}([s-H_1]^+,\alpha) \\ &\quad \geq w_{t+1}([s-\varepsilon-H_0]^+,\alpha) - w_{t+1}([s-\varepsilon-H_1]^+,\alpha), \end{aligned} \tag{39}$$

Then, we backward induce it at t. From (8), for $\alpha \in \mathcal{A}^{(0)}$, $\forall 0 < \varepsilon \leq \nu$, if $u_1, u_2, u_3, u_4 \in \tilde{\mathcal{U}}^{(\alpha)}$ are defined as optimal actions [12] for the following states e_1, e_2, e_3, e_4, namely:

$$\begin{aligned} w_t(e_1) &= w_t([s-H_0]^+,\alpha) \\ &= \min_{u\in\tilde{\mathcal{U}}^{(\alpha)}} \{\mu_t([s-H_0]^+,\alpha,u)\} \\ &= \mu_t([s-H_0]^+,\alpha,u_1) \end{aligned} \tag{40}$$

$$\begin{aligned} w_t(e_2) &= w_t([s-H_1]^+,\alpha) \\ &= \min_{u\in\tilde{\mathcal{U}}^{(\alpha)}} \{\mu_t([s-H_1]^+,\alpha,u)\} \\ &= \mu_t([s-H_1]^+,\alpha,u_2) \end{aligned} \tag{41}$$

$$\begin{aligned} w_t(e_3) &= w_t([s-\varepsilon-H_0]^+,\alpha) \\ &= \min_{u\in\tilde{\mathcal{U}}^{(\alpha)}} \{\mu_t([s-\varepsilon-H_0]^+,\alpha,u)\} \\ &= \mu_t([s-\varepsilon-H_0]^+,\alpha,u_3) \end{aligned} \tag{42}$$

$$\begin{aligned} w_t(e_4) &= w_t([s-\varepsilon-H_1]^+,\alpha) \\ &= \min_{u\in\tilde{\mathcal{U}}^{(\alpha)}} \{\mu_t([s-\varepsilon-H_1]^+,\alpha,u)\} \\ &= \mu_t([s-\varepsilon-H_1]^+,\alpha,u_4) \end{aligned} \tag{43}$$

We thus have (44) and (45) shown at the bottom of this page. In (45), the equalities by (34) and the inequalities are obtained by the hypothesis in (39). From (41) and (42), we have $B_1 \geq 0$ and $C_1 \geq 0$, respectively. With (45), we obtain

$$\begin{aligned} &w_t([s-H_0]^+,\alpha) - w_t([s-H_1]^+,\alpha) \\ &\quad - w_t([s-\varepsilon-H_0]^+,\alpha) + w_t([s-\varepsilon-H_1]^+,\alpha) \geq 0, \end{aligned} \tag{46}$$

which completes the proof of *Lemma 1* (a). *Lemma 1* (b) can be proved similarly. Then, with properties in *Lemma 1*, we prove the subadditive of $\mu_t(s,\alpha,u)$ in *Lemma 2*

Lemma 2: With $0 \leq H_0 \leq H_1$, $0 \leq H_2 \leq H_3$, $\forall 0 < \varepsilon \leq \nu$, $\chi(s)$ *is a convex and non-decreasing function in s, if*

(a) for $\alpha \in \mathcal{A}^{(0)}$, $\tilde{\mathcal{U}}^{(\alpha)} = \{0, 1\}$

$$\begin{aligned} &w_t([s-H_0]^+,\alpha) - w_t([s-H_1]^+,\alpha) \\ &\quad \geq w_t([s-\varepsilon-H_0]^+,\alpha) - w_t([s-\varepsilon-H_1]^+,\alpha), \end{aligned} \tag{47}$$

(b) for $\alpha \in \mathcal{A}^{(1)}$, $\tilde{\mathcal{U}}^{(\alpha)} = \{2, 3\}$

$$\begin{aligned} &w_t([s-H_2]^+,\alpha) - w_t([s-H_3]^+,\alpha) \\ &\quad \geq w_t([s-\varepsilon-H_2]^+,\alpha) - w_t([s-\varepsilon-H_3]^+,\alpha). \end{aligned} \tag{48}$$

then, $\mu_t(s,\alpha,u)$ *is subadditive on* $\mathcal{S} \times \tilde{\mathcal{U}}^{(\alpha)}$.

$$\begin{aligned} &w_t([s-H_0]^+,\alpha) - w_t([s-H_1]^+,\alpha) - w_t([s-\varepsilon-H_0]^+,\alpha) + w_t([s-\varepsilon-H_1]^+,\alpha) \\ &= \mu_t([s-H_0]^+,\alpha,u_1) - \mu_t([s-H_1]^+,\alpha,u_2) - \mu_t([s-\varepsilon-H_0]^+,\alpha,u_3) + \mu_t([s-\varepsilon-H_1]^+,\alpha,u_4) \\ &= \underbrace{\mu_t([s-H_0]^+,\alpha,u_1) - \mu_t([s-\varepsilon-H_0]^+,\alpha,u_1)}_{A_1} + \underbrace{\mu_t([s-\varepsilon-H_0]^+,\alpha,u_1) - \mu_t([s-\varepsilon-H_0]^+,\alpha,u_3)}_{B_1} \\ &\quad + \underbrace{\big(-\mu_t([s-H_1]^+,\alpha,u_2) + \mu_t([s-H_1]^+,\alpha,u_4)\big)}_{C_1} - \underbrace{\big(\mu_t([s-H_1]^+,\alpha,u_4) + \mu_t([s-\varepsilon-H_0]^+,\alpha,u_4)\big)}_{D_1} \\ &= A_1 + B_1 + C_1 - D_1 \end{aligned} \tag{44}$$

$$\begin{aligned} A_1 &= \left\{ \begin{array}{l} \sum\limits_{\alpha'\in\mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right) \left\{\delta(l=1)\left[w_{t+1}([s-H_0-H_1)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_1)]^+,\alpha')\right]\right. \\ \left. +(1-\delta(l=1))\left[w_{t+1}([s-2H_0)]^+,\alpha') - w_{t+1}([s-\varepsilon-2H_0)]^+,\alpha')\right]\right\} \\ + \sum\limits_{\alpha'\in\mathcal{A}^{(1)}} p\left(\alpha'|\alpha\right) \left\{\delta(l=1)\left[w_{t+1}([s-H_0-H_3)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_3)]^+,\alpha')\right]\right. \\ \left. +(1-\delta(l=1))\left[w_{t+1}([s-H_0-H_2)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_2)]^+,\alpha')\right]\right\} \end{array} \right. \\ &\geq \left\{ \begin{array}{l} \sum\limits_{\alpha'\in\mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right) \left[w_{t+1}([s-H_0-H_1)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_1)]^+,\alpha')\right] \\ + \sum\limits_{\alpha'\in\mathcal{A}^{(1)}} p\left(\alpha'|\alpha\right) \left[w_{t+1}([s-H_0-H_3)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_3)]^+,\alpha')\right] \end{array} \right. \\ &\geq \left\{ \begin{array}{l} \sum\limits_{\alpha'\in\mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right) \left\{\delta(l=1)\left[w_{t+1}([s-2H_1)]^+,\alpha') - w_{t+1}([s-\varepsilon-2H_1)]^+,\alpha')\right]\right. \\ \left. +(1-\delta(l=1))\left[w_{t+1}([s-H_0-H_1)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_1)]^+,\alpha')\right]\right\} \\ + \sum\limits_{\alpha'\in\mathcal{A}^{(1)}} p\left(\alpha'|\alpha\right) \left\{\delta(l=1)\left[w_{t+1}([s-H_1-H_3)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_1-H_3)]^+,\alpha')\right]\right. \\ \left. +(1-\delta(l=1))\left[w_{t+1}([s-H_1-H_2)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_1-H_2)]^+,\alpha')\right]\right\} \end{array} \right. \\ &= D_1 \end{aligned} \tag{45}$$

Proof: Let $s^+, s^- \in \mathcal{S}$, $u^+, u^- \in \mathcal{U}$, $l^+, l^- \in \{0, 1\}$, $\alpha \in \mathcal{A}$ and $t \in \mathcal{T}$, where $s^+ \geq s^-$ and $u^+ \geq u^-$, then we have (49) shown in bottom of this page. In (49), the equality is by (34). Since $p\left(\alpha'|\alpha\right) \geq 0, \forall \alpha, \alpha' \in \mathcal{A}, l^+, l^- \in \{0, 1\}$, we have $u^+, u^- \in \{0, 1\}$ for the first part (i) for $\alpha \in \mathcal{A}^{(0)}$, and so $\delta(l^- = 1) \leq \delta(l^+ = 1)$. With *Lemma 1* (35), the part (i) is non-negative. Similarly, part (ii) is for $\alpha \in \mathcal{A}^{(1)}$, and we have $u^+, u^- \in \{2, 3\}$, $\delta(l^- = 1) \leq \delta(l^+ = 1)$, and thus the part (ii) is also non-negative with *Lemma 1* (34). With *Definition 1* and non-negativity in (49), we can then conclude that $\mu_t(s, \alpha, u)$ is subadditive on $\mathcal{S} \times \tilde{\mathcal{U}}^{(\alpha)}$. □

$$\begin{aligned}
&\mu_t(s^+, \alpha, u^+) + \mu_t(s^-, \alpha, u^-) - \mu_t(s^+, \alpha, u^-) - \mu_t(s^-, \alpha, u^+) \\
&\quad = \underbrace{\sum_{\alpha' \in \mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right)\left(\delta(l^- = 1) - \delta(l^+ = 1)\right) \cdot \begin{bmatrix} w_{t+1}([s^+ - H_0]^+, \alpha) - w_{t+1}([s^+ - H_1]^+, \alpha) \\ -w_{t+1}([s^- - H_0]^+, \alpha) + w_{t+1}([s^- - H_1]^+, \alpha) \end{bmatrix}}_{(i)} \\
&\qquad + \underbrace{\sum_{\alpha' \in \mathcal{A}^{(1)}} p\left(\alpha'|\alpha\right)\left(\delta(l^- = 1) - \delta(l^+ = 1)\right) \cdot \begin{bmatrix} w_{t+1}([s^+ - H_2]^+, \alpha) - w_{t+1}([s^+ - H_3]^+, \alpha) \\ -w_{t+1}([s^- - H_2]^+, \alpha) + w_{t+1}([s^- - H_3]^+, \alpha) \end{bmatrix}}_{(ii)} \qquad (49)
\end{aligned}$$

APPENDIX D PROOF OF THEOREM 2

Proof: Since the cases for $\alpha \in \mathcal{A}^{(0)}$ and $\alpha \in \mathcal{A}^{(1)}$ have intercommunity in proving, we will give the proof of $\alpha \in \mathcal{A}^{(0)}$ in details in the following and the similar method can be applied for $\alpha \in \mathcal{A}^{(1)}$. We consider the case $\alpha \in \mathcal{A}^{(0)}$, $0 \leq H_0 \leq H_1$. Let $s^+, s^- \in \mathcal{S}$ and $t \in \mathcal{T}$, where $s^- = [s^+ - k\nu]^+$ and $k > 0$. With conditions of Theorem 2, by iteratively applying *Lemma 2*, we have

$$\begin{aligned}
&w_t([s^+ - H_0]^+, \alpha) - w_t([s^+ - H_1]^+, \alpha) \\
&\quad \geq w_t([s^+ - \nu - H_0]^+, \alpha) - w_t([s^+ - \nu - H_1]^+, \alpha) \\
&\quad \geq w_t([s^+ - k\nu - H_0]^+, \alpha) - w_t([s^+ - k\nu - H_1]^+, \alpha) \\
&\quad = 1 w_t([s^- - H_0]^+, \alpha) - w_t([s^- - H_1]^+, \alpha). \qquad (50)
\end{aligned}$$

Likewise, for $\alpha \in \mathcal{A}^{(1)}$, we could get

$$\begin{aligned}
&w_t([s^+ - H_2]^+, \alpha) - w_t([s^+ - H_3]^+, \alpha) \\
&\qquad \geq w_t([s^- - H_2]^+, \alpha) - w_t([s^- - H_3]^+, \alpha). \qquad (51)
\end{aligned}$$

Since $\mu_t(s, \alpha, u)$ is subadditive on $\mathcal{S} \times \tilde{\mathcal{U}}^{(\alpha)}$ proved in Appendix C. With the conclusion in [20], the optimal mode $\xi_t^*(s, \alpha)$ is a monotone non-decreasing function in s, and the optimal LTE usage function $l(s, t)$ is also non-decreasing by the mode selection function $\xi_t^*(s, \alpha)$ in (14). We define the threshold in s as $s^* = s_t{}^*(\alpha, t)$, where $l(s^*, t) = \delta\left[\xi_t^*\left(s^*(\alpha, t), \alpha\right)\right] = 1$, and $l(s^* - \nu, t) = \delta\left[\xi_t^*\left((s^*(\alpha, t) - \nu), \alpha\right)\right] = 0$, $\nu > 0$, which completes the proof of the threshold structure in data size s as (22). □

APPENDIX E INCREMENT OF $w_t(s, \alpha)$ IN t

Proof: The proof of the threshold structure in dimension t is based on that increment property of $w_t(s, \alpha)$ [25], which is illustrated in *Lemmas 3* and *Lemmas 4*.

Lemma 3: With $0 \leq H_0 \leq H_1$, $0 \leq H_2 \leq H_3$, $\forall 0 < \varepsilon \leq \nu$, *if* $\chi(s)$ *is a convex and non-decreasing function in s, then we have*

(a) For $\alpha \in \mathcal{A}^{(0)}$

$$\begin{aligned}
&w_{T+1}([s - H_0]^+, \alpha) - w_{T+1}([s - H_1]^+, \alpha) \\
&\quad \geq w_T([s - H_0]^+, \alpha) - w_T([s - H_1]^+, \alpha), \qquad (52)
\end{aligned}$$

(b) For $\alpha \in \mathcal{A}^{(1)}$

$$\begin{aligned}
&w_{T+1}([s - H_2]^+, \alpha) - w_{T+1}([s - H_3]^+, \alpha) \\
&\quad \geq w_T([s - H_2]^+, \alpha) - w_T([s - H_3]^+, \alpha). \qquad (53)
\end{aligned}$$

Proof: First, we prove the *Lemma 3* (a). By (10), we have

$$\begin{aligned}
\mathit{Left} &= w_{T+1}([s - H_0]^+, \alpha) - w_{T+1}([s - H_1]^+, \alpha) \\
&= \chi([s - H_0]^+) - \chi([s - H_1]^+). \qquad (54)
\end{aligned}$$

$$\begin{aligned}
\mathit{Right} &= w_T([s - H_0]^+, \alpha) - w_T([s - H_1]^+, \alpha) \\
&= \min\{\mu_T([s - H_0]^+, \alpha, 0), \mu_T([s - H_0]^+, \alpha, 1)\} - \min\{\mu_T([s - H_1]^+, \alpha, 0), \mu_T([s - H_1]^+, \alpha, 1)\} \\
&= \min\left\{\sum_{\alpha' \in \mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right) w_{T+1}([s - 2H_0]^+, \alpha'), \gamma_1 + \sum_{\alpha' \in \mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right) w_{T+1}([s - H_0 - H_1]^+, \alpha')\right\} \\
&\quad - \min\left\{\sum_{\alpha' \in \mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right) w_{T+1}([s - H_1 - H_0]^+, \alpha'),\ \gamma_1 + \sum_{\alpha' \in \mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right) w_{T+1}([s - 2H_1]^+, \alpha')\right\} \\
&= \min\left\{\chi([s - 2H_0]^+), \gamma_1 + \chi([s - H_0 - H_1]^+)\right\} - \min\left\{\chi([s - H_1 - H_0]^+), \gamma_1 + \chi([s - 2H_1]^+)\right\} \qquad (55)
\end{aligned}$$

Next, in (55), as shown at the bottom of the previous page, the qualities are due to (8), (14), and (11) respectively. We divide the two cases from the last equation in (55).

Case I: if $\chi([s-2H_0]^+) < \gamma_1 + \chi([s-H_0-H_1]^+)$, then

$$\begin{aligned}\gamma_1 &> \chi([s-2H_0]^+) - \chi([s-H_1-H_0]^+)\\ &\geq \chi([s-H_0-H_1]^+) - \chi([s-2H_1]^+), \qquad (56)\end{aligned}$$

where the second inequality is by that $\chi(s)$ is a convex and non-decreasing with s. Thus, we get $\gamma_1 + \chi([s-2H_1]^+) \geq \chi([s-H_0-H_1]^+)$. As a result, we have

$$\begin{aligned}Right &= \chi([s-2H_0]^+) - \chi([s-H_0-H_1]^+)\\ &\leq \chi([s-H_0]^+) - \chi([s-H_1]^+) = Left. \qquad (57)\end{aligned}$$

Case II: $\chi([s-2H_0]^+) \leq \gamma_1 + \chi([s-H_0-H_1]^+)$, then

$$\begin{aligned}Right &= \gamma_1 + \chi([s-H_0-H_1]^+)\\ &\quad - \min\left\{\chi([s-H_1-H_0]^+), \gamma_1 + \chi([s-2H_1]^+)\right\}. \qquad (58)\end{aligned}$$

The Case II could be further divided into two subcases:

Subcase II-(a): if $\chi([s-H_1-H_0]^+) \leq \gamma_1 + \chi([s-2H_1]^+$

$$\begin{aligned}Right &= \gamma_1 + \chi([s-H_0-H_1]^+) - \chi([s-H_1-H_0]^+)\\ &\leq \chi([s-2H_0]^+) - \chi([s-H_1-H_0]^+)\\ &\leq \chi([s-H_0]^+) - \chi([s-H_1]^+) = Left, \qquad (59)\end{aligned}$$

where the first inequality is by the condition in Case II, and the second inequality is due to $\chi(s)$ is convex and non-decreasing.

Subcase II-(b): if $\chi([s-H_1-H_0]^+) > \gamma_1 + \chi([s-2H_1]^+$

$$\begin{aligned}Right &= \chi([s-H_0-H_1]^+) - \chi([s-2H_1]^+)\\ &\leq \chi([s-2H_0]^+) - \chi([s-H_1-H_0]^+)\\ &\leq \chi([s-H_0]^+) - \chi([s-H_1]^+) = Left, \qquad (60)\end{aligned}$$

where the inequality is due to the fact that $\chi(s)$ is a convex and non-decreasing function in s. Combining theses cases, we conclude $Left \geq Right$, which completes the proof of *Lemma 3* (a). Similar approaches could be applied to prove *Lemma 3* (b).

Lemma 4: if $\chi(s)$ is a convex and non-decreasing function in s, we have

(a) for $\alpha \in \mathcal{A}^{(0)}$

$$\begin{aligned}&w_{t+1}([s-H_0]^+,\alpha) - w_{t+1}([s-H_1]^+,\alpha)\\ &\quad \geq w_t([s-H_0]^+,\alpha) - w_t([s-H_1]^+,\alpha), \qquad (61)\end{aligned}$$

(b) for $\alpha \in \mathcal{A}^{(1)}$

$$\begin{aligned}&w_{t+1}([s-H_2]^+,\alpha) - w_{t+1}([s-H_3]^+,\alpha)\\ &\quad \geq w_t([s-H_2]^+,\alpha) - w_t([s-H_3]^+,\alpha). \qquad (62)\end{aligned}$$

Proof: Similar to Appendix C, we prove it (a) by induction. First, from *Lemma 3*, we have the (61) held for $t = T$. Assume that for given $t \in \mathcal{T}$, we have

$$\begin{aligned}&w_{t+2}([s-H_0]^+,\alpha) - w_{t+2}([s-H_1]^+,\alpha)\\ &\quad \geq w_{t+1}([s-H_0]^+,\alpha) - w_{t+1}([s-H_1]^+,\alpha). \qquad (63)\end{aligned}$$

Suppose $\alpha \in \mathcal{A}^{(0)}$, $\forall 0 < \varepsilon \leq \nu$, if $u_5, u_6, u_7, u_8 \in \tilde{\mathcal{U}}^{(\alpha)}$ are optimal modes for the following states e_5, e_6, e_7, e_8, namely:

$$\begin{aligned}w_{t+1}(e_5) &= w_{t+1}([s-H_0]^+,\alpha)\\ &= \min_{u\in\tilde{\mathcal{U}}^{(\alpha)}} \{\mu_{t+1}([s-H_0]^+,\alpha,u)\}\\ &= \mu_{t+1}([s-H_0]^+,\alpha,u_5) \qquad (64)\\ w_{t+1}(e_6) &= w_{t+1}([s-H_1]^+,\alpha)\\ &= \min_{u\in\tilde{\mathcal{U}}^{(\alpha)}} \{\mu_{t+1}([s-H_1]^+,\alpha,u)\}\\ &= \mu_{t+1}([s-H_1]^+,\alpha,u_6)) \qquad (65)\\ w_{t+1}(e_7) &= w_{t+1}([s-\varepsilon-H_0]^+,\alpha)\\ &= \min_{u\in\tilde{\mathcal{U}}^{(\alpha)}} \{\mu_{t+1}([s-\varepsilon-H_0]^+,\alpha,u)\}\\ &= \mu_{t+1}([s-\varepsilon-H_0]^+,\alpha,u_7) \qquad (66)\\ w_{t+1}(e_8) &= w_{t+1}([s-\varepsilon-H_1]^+,\alpha)\\ &= \min_{u\in\tilde{\mathcal{U}}^{(\alpha)}} \{\mu_{t+1}([s-\varepsilon-H_1]^+,\alpha,u)\}\\ &= \mu_{t+1}([s-\varepsilon-H_1]^+,\alpha,u_8) \qquad (67)\end{aligned}$$

We thus have (68) as shown at the top of the next page. Further, for $\forall\varepsilon > 0$, we have (69), as shown at the top of the next page, where the two equalities are obtained by using (34) and the two inequalities are due to the induction hypothesis in (61) and (62). From (65) and (66), we have $B_2 > 0$ and $C_2 > 0$, respectively. Overall, from (68), we obtain

$$\begin{aligned}&w_{t+1}([s-H_0]^+,\alpha) - w_{t+1}([s-H_1]^+,\alpha)\\ &\quad \geq w_t([s-H_0]^+,\alpha) - w_t([s-H_1]^+,\alpha). \qquad (70)\end{aligned}$$

Lemma 4 (b) is proved with similar way. □

APPENDIX F
PROOF OF THEOREM 3

Proof: (a) For $\alpha \in \mathcal{A}^{(0)}$, there exists $t \in \mathcal{T}$ that satisfies $\mu_t(s,\alpha,1) \leq \mu_t(s,\alpha,0)$. Thus, we have $l(s,t) = 1$ and $\xi_t^*(s,\alpha) = 1$. Then,

$$\begin{aligned}\gamma_1 &< \sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right)\\ &\quad \left[w_{t+1}([s-H_0)]^+,\alpha')w_{t+1}([s-H_1)]^+,\alpha')-\right]\\ &\leq \sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right)\\ &\quad \times\left[w_{t+2}([s-H_0)]^+,\alpha') - w_{t+2}([s-H_1)]^+,\alpha')\right]\\ &= \mu_{t+1}(s,\alpha,0) - \mu_{t+1}(s,\alpha,1) + \gamma_1, \qquad (71)\end{aligned}$$

where the first inequality is by (14), and the second inequality is by *Lemma 4*, which implies $\mu_{t+1}(s,\alpha,1) \leq \mu_{t+1}(s,\alpha,0)$. Therefore, we have $l(s,t+1) = 1$ and $\xi_{t+1}^*(s,\alpha)$= 1

$$
\begin{aligned}
& w_{t+1}([s-H_0]^+,\alpha) - w_{t+1}([s-H_1]^+,\alpha) - w_{t+1}([s-\varepsilon-H_0]^+,\alpha) + w_{t+1}([s-\varepsilon-H_1]^+,\alpha)\\
&\quad = \mu_{t+1}([s-H_0]^+,\alpha,u_5) - \mu_{t+1}([s-H_1]^+,\alpha,u_6) - \mu_{t+1}([s-\varepsilon-H_0]^+,\alpha,u_7) + \mu_{t+1}([s-\varepsilon-H_1]^+,\alpha,u_8)\\
&\quad = \underbrace{\mu_{t+1}([s-H_0]^+,\alpha,u_5) - \mu_{t+1}([s-\varepsilon-H_0]^+,\alpha,u_5)}_{A_2} + \underbrace{\mu_{t+1}([s-\varepsilon-H_0]^+,\alpha,u_5) - \mu_{t+1}([s-\varepsilon-H_0]^+,\alpha,u_7)}_{B_2}\\
&\qquad + \underbrace{\big(-\mu_{t+1}([s-H_1]^+,\alpha,u_6) + \mu_{t+1}([s-H_1]^+,\alpha,u_8)\big)}_{C_2} - \underbrace{\big(\mu_{t+1}([s-H_1]^+,\alpha,u_8) + \mu_{t+1}([s-\nu-H_1]^+,\alpha,u_8)\big)}_{D_2}\\
&\quad = A_2 + B_2 + C_2 - D_2
\end{aligned} \tag{68}
$$

$$
\begin{aligned}
A_2 &= \begin{cases} \sum_{\alpha'\in\mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right)\left\{\delta(l=1)\left[w_{t+2}([s-H_0-H_1)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_1)]^+,\alpha')\right]\right.\\
\left. + (1-\delta(l=1))\left[w_{t+2}([s-2H_0)]^+,\alpha') - w_{t+1}([s-\varepsilon-2H_0)]^+,\alpha')\right]\right\}\\
+ \sum_{\alpha'\in\mathcal{A}^{(1)}} p\left(\alpha'|\alpha\right)\left\{\delta(l=1)\left[w_{t+2}([s-H_0-H_3)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_3)]^+,\alpha')\right]\right.\\
\left. + (1-\delta(l=1))\left[w_{t+2}([s-H_0-H_2)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_2)]^+,\alpha')\right]\right\}
\end{cases}\\
&\geq \begin{cases} \sum_{\alpha'\in\mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right) w_{t+2}([s-H_0-H_1)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_1)]^+,\alpha')\\
+ \sum_{\alpha'\in\mathcal{A}^{(1)}} p\left(\alpha'|\alpha\right) w_{t+2}([s-H_0-H_3)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_0-H_3)]^+,\alpha')
\end{cases}\\
&\geq \begin{cases} \sum_{\alpha'\in\mathcal{A}^{(0)}} p\left(\alpha'|\alpha\right)\ \delta(l=1)\left[w_{t+2}([s-2H_1)]^+,\alpha') - w_{t+1}([s-\varepsilon-2H_1)]^+,\alpha')\right]\}\\
+ (1-\delta(l=1))\left[w_{t+2}([s-H_1-H_0)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_1-H_0)]^+,\alpha')\right]\}\\
+ \sum_{\alpha'\in\mathcal{A}^{(1)}} p\left(\alpha'|\alpha\right)\left\{\delta(l=1)\left[w_{t+2}([s-H_1-H_3)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_1-H_3)]^+,\alpha')\right]\right.\\
\left. + (1-\delta(l=1))\left[w_{t+2}([s-H_1-H_2)]^+,\alpha') - w_{t+1}([s-\varepsilon-H_1-H_2)]^+,\alpha')\right]\right\}
\end{cases}\\
&= D_2
\end{aligned} \tag{69}
$$

(b) For $\alpha \in \mathcal{A}^{(1)}$, assume that $t \in \mathcal{T}$ satisfies that $\mu_t(s,\alpha,2) \leq \mu_t(s,\alpha,3)$, i.e., $\xi_t^*(s,\alpha) = 3$ and $l(s,t) = 1$.

$$
\begin{aligned}
\gamma_3 &< \sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right)\\
&\quad \times \left[w_{t+1}([s-H_2)]^+,\alpha') - w_{t+1}([s-H_3)]^+,\alpha')\right]\\
&\leq \sum_{\alpha'\in\mathcal{A}} p\left(\alpha'|\alpha\right)\\
&\quad \times \left[w_{t+2}([s-H_2)]^+,\alpha') - w_{t+2}([s-H_3)]^+,\alpha')\right]\\
&= \mu_{t+1}(s,\alpha,3) - \mu_{t+1}(s,\alpha,2) + \gamma_3.
\end{aligned} \tag{72}
$$

Then, we obtain $\mu_{t+1}(s,\alpha,3) \leq \mu_{t+1}(s,\alpha,2)$, and thus $\xi_{t+1}^*(s,\alpha)= 3$, $l(s,t+1) = 1$. To conclude, we show for $\forall\alpha \in \mathcal{A}$, the LTE usage l is non-decreasing in t. Namely, the threshold of LTE usage in t exists as indicated in (25). □

APPENDIX G
PROOF OF THEOREM 4

Proof: (a) Let $\alpha \in \mathcal{A}$ and $t \in \mathcal{T}$ be given. By the definition of the threshold $s^*(\alpha,t)$ in (23) and (24), we have $l^*(s,t) = 0$ if $0 \leq s \leq s^*(\alpha,t)$. From the threshold structure in time t in (26) and (27), it implies that $l^*(s,t-1) = 0$ if $0 \leq s \leq s^*(\alpha,t)$. By the definition of threshold $s^*(\alpha,t-1)$ at time $t-1$, we can conclude that $s^*(\alpha,t-1) \geq s^*(\alpha,t)$.

(b) Let $\alpha \in \mathcal{A}$ and $s \in \mathcal{S}$ be given. By the definition of threshold $t^*(s,\alpha)$ in (26) and (27), we have $l^*(s,t) = 1$ when $t \geq t^*(s,\alpha)$, From threshold structure in s in (23) and (24), it implies that $l^*(s+\nu,t) = 1$ for $t \geq t^*(s,\alpha)$. By the definition of threshold, we conclude $t^*(s,\alpha) \geq t^*(s+\nu,\alpha)$. □

ACKNOWLEDGEMENT

This paper was presented in part at the Proc. 85th Vehicular Technology Conference, Sydney, Australia, in Jun. 2017 [1].

REFERENCES

[1] B. Liu, Q. Zhu, and H. Zhu, "Delay-aware LTE WLAN aggregation for 5G unlicensed spectrum usage," in *Proc. 85th Veh. Technol. Conf. (VTC Spring)*, Sydney, NSW, Australia, Jun. 2017, pp. 1–7.

[2] Cisco, Visual Networking Index, "Global mobile data traffic forecast update, 2016–2021," Cisco, San Jose, CA, USA, White Paper 1454457600805266, Feb. 2017.

[3] J. Huang, Y. Zhou, and X. Cong, "Game-theoretic power control mechanisms for Device-to-Device communications underlaying cellular system," *IEEE Trans. Veh. Technol.*, to be published.

[4] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can WiFi deliver?" *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 536–550, Apr. 2013.

[5] P. Nuggehalli, "LTE-WLAN aggregation [industry perspectives]," *IEEE Wireless Commun.*, vol. 23, no. 4, pp. 4–6, Aug. 2016.

[6] *New Work Item on Enhanced LWA*, document RP-160600, 3GPP RAN, Göteborg, Sweden, Mar. 2016.

[7] *M1, Nokia Announce Singapore's First Commercial Nationwide HetNet Rollout*, Nokia, Espoo, Finland, Mar. 2017.

[8] D. Suh, H. Ko, and S. Pack, "Efficiency analysis of WiFi offloading techniques," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3813–3817, May 2016.

[9] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proc. ACM MobiSys*, Jun. 2010, pp. 209–222.

[10] H. Deng and I.-H. Hou, "On the capacity-performance trade-off of online policy in delayed mobile offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 526–537, Jan. 2017.

[11] N. Wang and J. Wu, "Opportunistic WiFi offloading in a vehicular environment: Waiting or downloading now?" in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.

[12] M. H. Cheung and J. Huang, "DAWN: Delay-aware Wi-Fi offloading and network selection," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1214–1223, Jun. 2015.

[13] H. Ko, J. Lee, and S. Pack, "Performance optimization of delayed WiFi offloading in heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9436–9447, Oct. 2017.

[14] J. Lee, Y. Yi, S. Chong, and Y. Jin, "Economics of WiFi offloading: Trading delay for cellular capacity," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1540–1554, Mar. 2014.

[15] Q. Zhu *et al.*, "A digital polar transmitter with DC–DC converter supporting 256-QAM WLAN and 40-MHz LTE-A carrier aggregation," *IEEE J. Solid-State Circuits*, vol. 52, no. 5, pp. 1196–1209, Mar. 2017.

[16] Y. Ohta, N. Michiharu, S. Aikawa, and T. Ode, "Link layer structure for LTE-WLAN aggregation in LTE-advanced and 5G network," in *Proc. IEEE Conf. Standards Commun. Netw. (CSCN)*, Oct. 2015, pp. 83–88.

[17] S. Singh, S.-P. Yeh, N. Himayat, and S. Talwar, "Optimal traffic aggregation in multi-RAT heterogeneous wireless networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 626–631.

[18] S. Singh, M. Geraseminko, S.-P. Yeh, N. Himayat, and S. Talwar, "Proportional fair traffic splitting and aggregation in heterogeneous wireless networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1010–1013, Mar. 2016.

[19] D. López-Pérez *et al.*, "Long term evolution-wireless local area network aggregation flow control," *IEEE Access*, vol. 4, pp. 9860–9869, 2016.

[20] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY USA: Wiley, 2014.

[21] 5G Americas, "LTE aggregation unlicensed spectrum," Bellevue, WA, USA, White Paper, Nov. 2015.

[22] L. Zhang, W. Wu, and D. Wang, "Time dependent pricing in wireless data networks: Flat-rate vs. usage-based schemes," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Toronto, ON, Canada, May 2014, pp. 700–708.

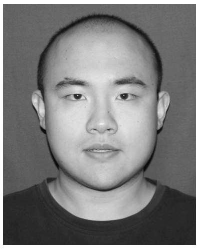
[23] J. Huang, C. Xing, Y. Qian, and Z. J. Haas, "Resource allocation for multicell device-to-device communications underlaying 5G networks: A game-theoretic mechanism with incomplete information," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2017.2765208.

[24] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "Incentivizing time-shifting of data: A survey of time-dependent pricing for Internet access," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 91–99, Nov. 2012.

[25] Y. Wu, Q. Yang, X. Liu, and K. S. Kwak, "Delay-constrained optimal transmission with proactive spectrum handoff in cognitive radio networks," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2767–2779, Jul. 2016.

[26] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Belmont, MA, USA: Athena Scientific, 2005.

[27] M. H. Ngo and V. Krishnamurthy, "Optimality of threshold policies for transmission scheduling in correlated fading channels," *IEEE Trans. Commun.*, vol. 57, no. 8, pp. 2474–2483, Aug. 2009.



BIN LIU (S'17) received the B.S. degree from the Changshu Institute of Technology, Suzhou, China, in 2015. He is currently pursuing the Ph.D. degree in communications and information engineering with the Nanjing University of Posts and Telecommunications, Nanjing, China. His current research interests include mobile data offloading, LTE-U, and unlicensed spectrum allocation in next generation wireless communication systems.

QI ZHU received the bachelor's and master's degrees in radio engineering from the Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China, in 1986 and 1989, respectively. She is currently a Professor with the School of Telecommunication and Information Engineering, NUPT. Her research interests include technology of next-generation communication, broadband wireless access, orthogonal frequency division multiplexing, channel and source coding, and dynamic allocation of radio resources.

HONGBO ZHU received the B.S. degree in communications engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982, and the Ph.D. degree in information and communications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 1996. He is currently a Professor and the Vice President of the Nanjing University of Posts and Telecommunications, Nanjing. He is also the Head of the Coordination Innovative Center of IoT Technology and Application (Jiangsu), which is the first governmental authorized Coordination Innovative Center of Internet of Things (IoT) in China. He also serves as a referee or expert in multiple national organizations and committees. He has authored and co-authored over 200 technical papers published in various journals and conferences. He is currently leading a big group and multiple funds on IoT and wireless communications with current focus on architecture and enabling technologies for IoT. His research interests include mobile communications, wireless communication theory, and electromagnetic compatibility.