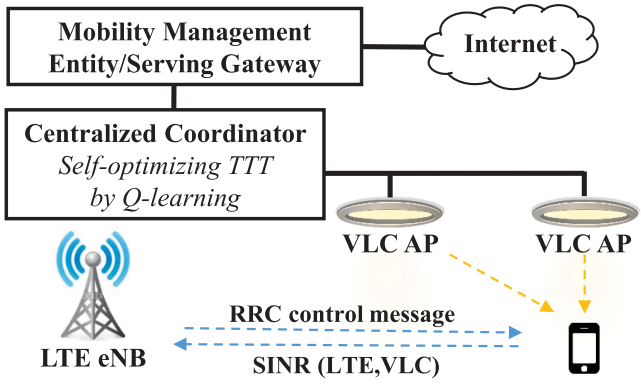
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Optimizing Handover Parameters by Q-Learning for Heterogeneous Radio-Optical Networks

Volume 12, Number 1, February 2020

Sihua Shao
Guanxiong Liu
Abdallah Khreishah
Moussa Ayyash
Hany Elgala
Thomas D. C. Little
Michael Rahaim



DOI: 10.1109/JPHOT.2019.2953863

Optimizing Handover Parameters by Q-Learning for Heterogeneous Radio-Optical Networks

Sihua Shao [1], Guanxiong Liu,[2] Abdallah Khreishah [2], Moussa Ayyash [3,4], Hany Elgala [5], Thomas D. C. Little [6], and Michael Rahaim[7]

[1]Department of Electrical Engineering, New Mexico Tech, Socorro, NM 87801 USA
[2]Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA
[3]Department of Computer and Mathematical Sciences, Lewis University, Romeoville, IL 60446 USA
[4]Department of Information Studies, Chicago State University, Chicago, IL 60628 USA
[5]Department of Electrical and Computer Engineering, SUNY at Albany, Albany, NY 12222 USA
[6]Department of Electrical and Computer Engineering, Boston University, MA 02215 USA
[7]Department of Engineering, University of Massachusetts Boston, MA 02125 USA

DOI:10.1109/JPHOT.2019.2953863

Manuscript received October 21, 2019; revised November 11, 2019; accepted November 13, 2019. Date of publication November 18, 2019; date of current version January 15, 2020. This work was supported in part by the National Science Foundation under Grants CNS-1617924, CNS-1617866, and CNS-1617645. Corresponding author: Sihua Shao (e-mail: sihua.shao@nmt.edu).

Abstract: Existing literature studying the access point (AP)-user association problem of heterogeneous radio-optical networks either investigates quasi-static network selection or only considers vertical handover (VHO) dwell time from optical to radio. The quasi-static assumption can result in outdated decisions for highly mobile scenarios. Solely focusing on the optical to radio handover ignores the importance of dwell time for VHO from radio to optical. In this paper, we propose a flexible and holistic framework, that runs a self-optimizing algorithm at the centralized coordinator (CC). This CC resides in the LTE eNodeB and controls the handover parameters of all the visible light communication (VLC) APs under the coverage of the LTE eNodeB. Based on Q-learning approach, the algorithm optimizes the time-to-trigger (TTT) values for VHO between LTE and VLC. Case studies are performed to validate the considerable gain in terms of average throughput by optimizing TTTs. We evaluate the impact of learning parameters on the optimal throughput and convergence speed through trace-driven simulations. The simulation results reveal that the Q-learning based algorithm improves the average throughput of mobile device by 25% when compared to the fixed TTT scheme. Furthermore, this algorithm is capable of self-optimizing handover parameters in an online manner.

Index Terms: Handover, Q-learning, heterogeneous network, visible light communication, parameter optimization.

1. Introduction

To meet the ultra-high speed requirement of the 5G network [1], frequency bands in the mmWave and THz range are being explored. These extremely high frequencies are expect to supplement the sub-6 GHz bands currently used in 4G. Heterogeneous network (HetNet) [2] integration of low-power small cells within high power macrocells is a low-cost and energy-efficient solution to

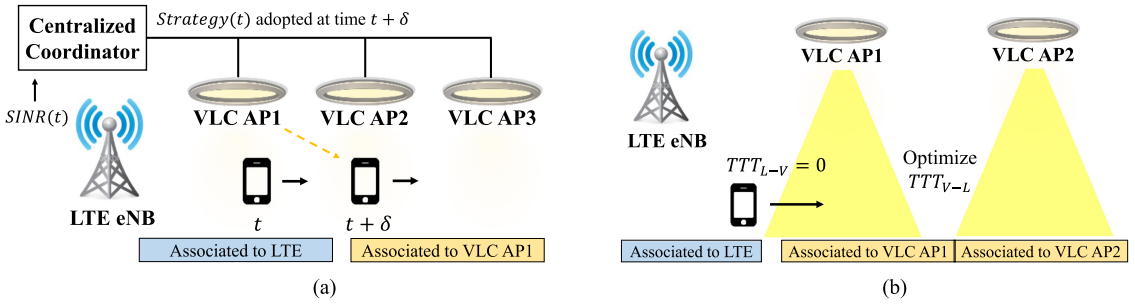


Fig. 1. (a) Quasi-static network selection. (b) Assumed immediate handover from RF to VLC.

satisfy the specifications of user experienced data rate, mobility and latency in 5G standards. Enjoying advantages such as dual-use nature, high energy efficiency, ubiquitous availability and no interference to radio frequency (RF)-based devices, optical wireless communication (OWC) technologies such as visible light communication (VLC) [3] are attractive options to offload data traffic from RF-based networks.

The quasi-static network model of the access point (AP)-user association problem of RF-VLC HetNets was investigated in [4]–[6], where the channel characteristics are assumed to be fixed in each coherent and equal-length time slot. As shown in Fig. 1(a), according to the signal-to-interference-plus-noise ratio (SINR) feedback at time t, the centralized coordinator (CC) generates the strategy based on some proposed algorithm. However, the generated AP-user association strategy starts playing an effect at time $t + \delta$. The literature based on quasi-static network model assumes the user equipment (UE) change their location or orientation to a negligible extent within time δ, which allows the CC to compute and generate the strategy. According to the experimental results (Fig. 4(a) in [7]), if we consider the SNR variation in a walking in-and-out scenario of a VLC cell, at the steepest region, it takes only 10 ms for the SNR to change by 10 dB. Therefore, it is highly possible that the strategy generated at time t will be outdated at time $t + \delta$ under mobile scenarios and will lead to significant deterioration of signal quality or even frequent disconnection.

The dwell time (i.e., the amount of time by which handover is delayed after condition is met) for vertical handover (VHO) from VLC to RF was investigated in [8]–[10], which assume immediate handover from RF to VLC whenever the condition is met. As shown in Fig. 1(b), the literature studying the VHO between RF (e.g., LTE eNodeB) and VLC always assume the value of the time-to-trigger parameter TTT_{L-V} from LTE to VLC is equal to zero, which indicates that the immediate handover will occur once the handover condition (i.e., events B1 or B2 in [11]) is met. However, considering the quality of experience (QoE) while the UE is crossing the VLC cell, zero-value TTT_{L-V} may not be the best option and sometimes might even lead to intermittent disconnection. Later on in Section 5 we will show based on thorough case studies that the value of TTT_{L-V} has an important impact on the average throughput which also depends on the congestion level in RF network and the user movement speed.

To properly resolve the AP-user association problem in RF-VLC HetNets, we aim to find a flexible and holistic solution which adaptively varies the handover parameters in a self-organizing manner and enables the capability of context-aware (e.g., cell load, user profile, etc.) decision making. To accomplish the goal, we propose a framework (Section 3) incorporating LTE and VLC networks by leveraging the LTE uplink as the common channel for SINR feedback. In a CC, the TTT values (i.e., both TTT_{L-V} and TTT_{V-L}) controlling the VHO between LTE and VLC are optimized by our designed Q-learning based algorithm. The algorithm incorporates historical SINR feedback in order to maximize average throughput. This is particularly impactful for the mobile user scenario.

We review the related works in Section 2 that rely on quasi-static network model to optimize the network selection in each coherent state or assume immediate handover from RF to VLC once the VHO condition is met. In Section 4, we introduce the basic principles of Q-learning and

demonstrate the Q-learning based algorithm for the self-optimization of handover parameters in LTE-VLC HetNets. Case studies (Section 5) are performed to evaluate the impact of different values of handover parameters (i.e., TTT_{L-V} and TTT_{V-L}) on the average throughput under different system settings. In Section 6, based on real-subject dataset [12], [13], we verify the repetitive nature of user translational motion over a 24 hour cycle. Simulations based on real measurements (Section 7) are conducted to evaluate the appropriate strategy of selecting the Q-learning based algorithm parameters. Simulation results reveal that larger state space leads to higher achievable system performance; but also requires increasing the algorithm's convergence time. We summarize our work in Section 8.

In summary, we make five key contributions:

- We propose a flexible and holistic framework to coexist the VLC and LTE networks and enable the optimization of handover parameters in a self-organized manner.
- We design a Q-learning based algorithm to optimize a sequence of (TTT_{L-V}, TTT_{V-L}) in order to maximize the average throughput or mobile UEs.
- We perform thorough case studies to validate the necessity of optimizing both TTT_{L-V} and TTT_{V-L}.
- We verify the repetitiveness of user translational movement pattern over a day based on the evaluation of real-subject dataset.
- Extensive simulations are conducted to evaluate the impact of different learning algorithm parameters (e.g., trade-off factor ϵ, state space, number of time slots) on the converged throughput and convergence speed. Simulation results show that compared to the fixed TTT scheme, the Q-learning based algorithm improves the average throughput by 25% for mobile users. As the number of time slots increases, the converged throughput of Q-learning gets closer to the optimal performance at the cost of convergence speed.

2. Related Works

Quasi-static network selection. The work in [4] uses knowledge transfer to reduce the convergence time of Q-learning at the cost of the memory storage for the context-decision pairs. The scalability of the knowledge transfer approach proposed in [4] is limited. As the resolution of the context (i.e., traffic type, user location, time) increases, the consumed memory and the decision searching time increase significantly. The work in [5] divides the dynamic system into quasi-static states and the channel state information (CSI) is assumed to be fixed for each state. If one user is assigned to two different APs in two consecutive states, it means that the handover occurs. The handover overhead follows independent and identical Poisson distribution. In [6], UEs select APs based on the prediction of user trajectory and APs perform load balancing by a dynamic load graph. The proposed algorithm in [6] integrates "user choosing AP" and "AP selecting user" to improve the preferences such as system throughput, load fairness and handover overhead. For highly mobile scenarios, the decision made based on static network pattern will generally be outdated which may lead to outage problem and worse system performance than that of simply connecting the UEs to the AP with the highest SNR. Controlling the handover parameters of each cell in a centralized manner while leaving the AP-user association to some handover trigger conditions (e.g., SNR of server becomes worse than threshold, SNR of neighbor becomes better than threshold, etc.) is more suitable for mobile scenarios.

Immediate handover from RF to VLC. In [8], the time interval of crossing the boundary between VLC and WLAN networks, which indicates the frequency of handover event, is utilized to determine the length of dwell time for handover from VLC to WLAN. Shorter interval leads to longer dwell time and vice versa. The work in [9] investigates the decision-making problem of performing VHO when VLC channel condition is not satisfactory. The work employs analytic hierarchy process and cooperative game in order to handle the multi-attribute decision-making process and fit various traffic types. Handover procedures in RF-VLC HetNets are elaborated in [10] which includes the VLC frame design, channel access in asymmetric RF-VLC system, VHO, horizontal handover, multi-user scenario and mobility issues. However, to the best of our knowledge, all the literature

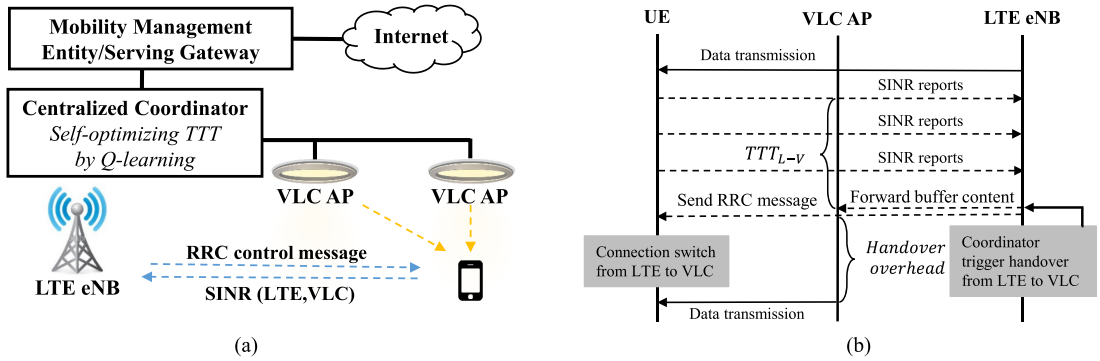


Fig. 2. (a) LTE-VLC System Architecture. (b) Handover procedures from LTE to VLC.

studying VHO between RF and VLC assume the UE immediately attempts to connect to a VLC AP whenever it is within the coverage. Although opportunistically offloading data traffic from RF to VLC mitigates the congestion in RF networks, the resulted user experienced data rate and latency may be unacceptable especially for a UE crossing the small-range VLC APs at a fast moving speed.

Other related works. The work in [14] investigates the intermittent line-of-sight (LOS) blockage of VLC channels. The VHO problem is formulated as a Markov decision process and the objective focuses on the trade-off between energy consumption and delay requirement. In [15], a proof-of-concept hybrid VLC-RF system with handover enabled is implemented based on USRPs, GNU Radio and Linux bonding driver. The experimental results reveal that the handover overhead (i.e., from the time instant when the UE disconnects from the current serving cell to the time instant when the UE finishes the association process with the new cell) is over 800 ms. The handover overhead is considerable such that the UE may be entirely held in the handover process while it is crossing a VLC cell.

3. System Model

To optimize the handover parameters of LTE-VLC HetNets in a self-organized manner with context awareness, we propose a framework that is running a self-optimizing algorithm in a CC which may reside in the LTE eNodeB to control the handover parameters of VLC APs located under the coverage of the LTE eNodeB (Fig. 2(a)). We consider a control and user plane framework similar to that proposed in [16]. However, we aim to self-optimize the handover parameters instead of heuristically adjusting them. As shown in Fig. 2(b), the UEs report the SINR measurements of the potential serving LTE/VLC APs periodically to the CC through the LTE uplink control signaling. The coordinator initiates the handover process by sending a Radio Resource Control (RRC) message to the UE.

WiFi is an alternative radio access technology (RAT) that can be integrated into our proposed framework. One of the two basic functions in WiFi MAC layer is the Point Coordination Function (PCF) [17], which provides contention-free services by making the AP as a polling master for granting permissions to UEs in the range to transmit. PCF allows the CC to initiate the VHO in a timely manner (i.e., fitting well in our proposed framework) since the channel access delay induced by carrier-sense multiple access with collision avoidance (CSMA/CA) in distributed coordination function (DCF) can be avoided. However, in this paper, we focus on the discussion based on LTE-VLC HetNets and consider the integration of WiFi, especially the commonly adopted DCF mode, as our future work.

Typically, to design a handover policy, time-to-trigger (TTT) and hysteresis margin (HM) are set to appropriate values to lower the drop rate and the ping-pong rate [18]. However, since the line-of-sight (LOS) signal is dominant in VLC [19], the value of SNR either increases or decreases

monotonically when the UE crosses the VLC cell edge. Therefore, the value of HM, which is leveraged to alleviate the ping-pong effect, is set to zero for the VHO between RF and VLC. Based on the periodical SINR measurements report, the CC residing in the LTE eNodeB estimates the achieved average throughput and optimizes the values of TTT_{L-V} and TTT_{V-L} for the VHO from LTE to VLC and from VLC to LTE, respectively.

In this work, we assume that the distribution of user translational movement pattern over a day (i.e., 24 hours) is statistically repetitive. The assumption is rational due to the regular user behaviors for daily working and living, for example, the entrance/exit of an office building will be generally crowded during the time when people are going to work and getting off work, and many students will cross the hallway leading to the dining hall at a relatively low speed during the lunch hours. The assumption is also validated in Section 6 based on real-subject dataset. Therefore, we divide one day into multiple equal-length time slots and the values of TTT_{L-V} and TTT_{V-L} are fixed in each time slot. Since the handover process is controlled by the CC instead of individual users, both the values of the system-centric TTT_{L-V} and TTT_{V-L} in each time slot are the same for all the mobile users. However, A Q-learning based algorithm is designed to maximize the average throughput through learning the optimal sequence of TTTs which controls the handover behavior in each time slot.

4. Q-Learning Based Handover Parameter Self-Optimization

Machine learning can be widely used in modeling diverse technical problems of next-generation wireless networks [20], such as large-scale MIMOs, device-to-device (D2D) networks, HetNets constituted by femtocells and small cells, and so on. The family of learning techniques consists of supervised learning (e.g., regression model), unsupervised learning (e.g., K-means clustering) and reinforcement learning (e.g., Q-learning). In our work, we focus on the model-free and off-policy Q-learning based algorithm which is suitable for handling the dynamic environment of HetNets.

4.1 Q-Learning Algorithm

Q-learning is an off-policy temporal-difference control algorithm in reinforcement learning [21]. In Q-learning, an agent tries to discover the optimal policy according to the historical interactions with the environment. The experience $(s_t, a_t, r_{t+1}, s_{t+1})$ (where s_t is the current state, a_t is the action taken in the current state, r_{t+1} is the reward received by taking action a_t in state s_t and s_{t+1} is the next state after taking action a_t in state s_t) is utilized to update a lookup table, called Q-table. A converged Q-table contains the maximum expected total future rewards $Q(s_t, a_t)$ corresponding to each action a_t in state s_t. The highest score (i.e., the value of $Q(s_t, a_t)$) in each column of the Q-table (columns are actions and rows are states) indicates the best action in each state. The score $Q(s_t, a_t)$ corresponding to state s_t and action a_t is updated by

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)], \tag{1}$$

where $\alpha \in [0, 1]$ is the learning rate, γ is the discount factor and $\max_a Q(s_{t+1}, a)$ represents the maximum expected future reward given the state s_{t+1} and all possible actions in the state.

Although the target policy of Q-learning algorithm is to select the action corresponding the highest score (i.e., exploitation) in each state, the Q-table training process needs a different behavior policy to select the action randomly (i.e., exploration) in order to update the $Q(s_t, a_t)$ that may lead to global optimum. ϵ-greedy strategy is used to handle the trade-off between exploration and exploitation during the training of Q-table. In each state, the action corresponding to the highest score is selected with probability $1 - \epsilon$ and otherwise the action is selected randomly. The value of ϵ will be decreased along with the increase of the number of episodes N_{ep} [22],

$$\epsilon = \frac{\epsilon_{init}}{(1 + N_{ep}/a)^b}, \tag{2}$$

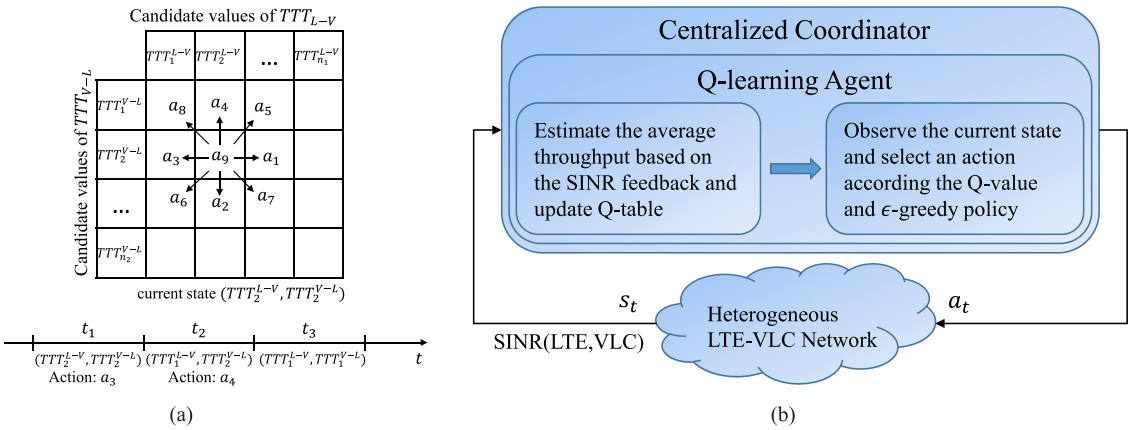


Fig. 3. (a) State and action space.(b) Diagram of Q-learning based algorithm.

where ϵ_{init} is the initial value of ϵ and the decreasing speed of ϵ is influenced by the values of a and b. The value of ϵ will be reset to ϵ_{init} when sudden degradation of performance occurs in order to adapt to the changes of repetitive user movement patterns.

4.2 Handover Parameter Self-Optimization

One important feature of Q-learning is that the next state depends on the current state and the action taken at the current state. Therefore, the state space is defined as the set of (TTT_{L-V}, TTT_{V-L}) and the size of the state space depends on the number of candidate values of TTT_{L-V} and TTT_{V-L}. Typically, the daily behavior of users (e.g., moving speed, trajectory, spatial density, etc) follows certain regular pattern, especially for indoor environments. An assumption is thereby made that the distribution of user translational movement pattern under each VLC cell over a day (i.e., 24 hours) is repetitive. The assumption indicates that if we divide one day into t_s equal-length time slots, there will be an optimal pair of (TTT_{L-V}, TTT_{V-L}) in each time slot such that the average throughput for the users crossing the VLC cells is maximized. Note that the major beneficiaries of optimizing TTTs are mobile users since the handover overhead becomes negligible when we consider the long-term average throughput of static users under the VLC cells. The action space is defined as the changes to the values of TTT_{L-V} and TTT_{V-L} in each state. In order to enable refined optimization of the handover parameters and avoid suddenly large changes, the action taken in each state increases or decreases the values of TTT_{L-V} and TTT_{V-L} by a single level. For example, in Fig. 3(a), given the current state as $(TTT_2^{L-V}, TTT_2^{V-L})$ at time slot t_1, the possible next states are restricted to the states surrounding the current one. In addition, keeping the values of TTT_{L-V} and TTT_{V-L} as they are in the current state is also considered as a candidate action such that the optimal TTT_{L-V} and TTT_{V-L} will not jump back and forth between two states. When action a_3 is taken at time slot t_1, the state at time slot t_2 will be $(TTT_1^{L-V}, TTT_2^{V-L})$. When action a_4 is taken at time slot t_2, the state at time slot t_3 will be $(TTT_1^{L-V}, TTT_1^{V-L})$. The reward r_{t+1} received by taking action a_t in state s_t is calculated by averaging the throughput (i.e., $\frac{\sum_i T(SINR)}{i}$ where $T(SINR)$ is the throughput estimation and i is the number of SINR measurement feedback) during time slot $t+1$ within which the selected values of TTT_{L-V} and TTT_{V-L} for state s_{t+1} are fixed for all the users. As shown in Fig. 3(b), the Q-learning agent estimates the reward based on the SINR feedback and updates the Q-table accordingly. The action in each state is selected based on the Q-value of each state-action pair and ϵ-greedy policy. Note that the initial state $s_1 = (TTT_{L-V}, TTT_{V-L})$ is reset to the same value s_{init} in each episode (i.e., each day). The state space, action space, reward function and exploration process are summarized as follows:

- *State space*: State space consists of the combinations of TTT_{L-V} and TTT_{V-L}. Assume the number of candidate TTT_{L-V} and TTT_{V-L} values are n_1 and n_2, respectively, the total number of states is $n_1 \times n_2$ (Fig. 3(a)).
- *Action space*: There are totally nine possible actions for each state (TTT_{L-V}, TTT_{V-L}) except those states where the value of TTT_{L-V} or TTT_{V-L} cannot further increase or decrease.
- *Reward function*: Assume there are totally k SINR feedback values within one time slot, the reward r_{t+1} is the average throughput T_{avg} in time slot $t+1$.

$$T_{avg} = \frac{1}{k}\left(\sum_i T_{LTE}(SINR_i) + \sum_j T_{VLC}(SINR_j)\right), \tag{3}$$

where $T_{LTE}(SINR_i)$ and $T_{VLC}(SINR_j)$ are the throughput estimations of LTE and VLC channels given the pre-determined bandwidth, respectively. The value of the non-zero throughput is estimated by Shannon equation in the simulations. Note that $i + j \leq k$ since if the UE is in the process of handover, the throughput estimation is considered as zero.

Algorithm 1: Q-Learning Based Algorithm for Optimizing Handover Parameters.

Initialize: Q-table, state s_{init}, ϵ_{init}, ϵ control factors a and b, number of time slots t_s, learning rate α, discount factor γ and candidate values of TTT_{L-V} and TTT_{V-L}.

Loop for each episode:
Initialize $t = 1$ and $s_t = s_{init}$;
while $t < t_s$ **do**
  Select action a_t according to ϵ-greedy policy;
  Take action a_t and observe r_{t+1} and s_{t+1};
  $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$;
  $s_t \leftarrow s_{t+1}$;
end while

- *Exploration process*: The Q-table is initialized with random values in the range $[-1, 1]$. The initial value of ϵ is set to ϵ_{init} and keeps decreasing along with the increase of the number of episodes as in (2). At the beginning of each episode, the state is reset to a pre-determined initial state s_{init}. For example, in the early morning, typically the utilization of LTE spectrum is low such that the initial value of TTT_{L-V} should be relatively large. The speculation is verified by case studies in Section 5. In each state, the action is selected according to the ϵ-greedy policy and the Q-table is updated by (1). Each episode ends at state s_{t_s}, where t_s is the last time slot during a day. The Q-learning based algorithm for optimizing handover parameters is summarized in Algorithm 1.

5. Case Study

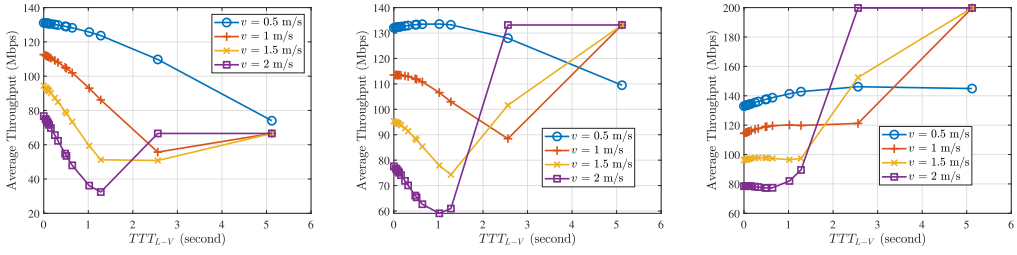
In this section, case studies[1] are performed in MATLAB R2018 under two different system settings: 1) one VLC AP and 2) two VLC APs, to reveal the limitations of the state-of-the-art VHO schemes (i.e., $TTT_{L-V} = 0$) and the considerable throughput gain from jointly optimizing the values of TTT_{L-V} and TTT_{V-L}. According to the specifications in [23], 16 candidate values (i.e., 0, 0.04, 0.064, 0.08, 0.1, 0.128, 0.16, 0.256, 0.32, 0.48, 0.512, 0.64, 1.024, 1.28, 2.56 and 5.12 seconds) are selected for both TTT_{L-V} and TTT_{V-L}. Therefore, the size of the state space is 256.

Based on the real measurements in [24], we set the system parameters to the values summarized in Table 1. The handover execution time is set to 800 ms, within which the throughput is zero. The SINR feedback is reported by the UEs through the LTE uplink every 10 ms. The semi-angle at half power and emitted optical power of each VLC AP are set to 60° and 1 W, respectively.

[1] https://github.com/Sihua-Shao/Handover

TABLE 1
System Parameters

Handover overhead	800 ms
SINR feedback interval	10 ms
User movement speed	0.5/1/1.5/2 m/s
Distance between adjacent VLC APs	3 m
Semi-angle at half power of VLC AP	60°
Total emitted optical power of one VLC AP	1 W
Sensing area of PD	10^{-4} m^2
FOV of VLC receiver	90°
O/E efficiency of VLC receiver	0.53 A/W
VLC channel bandwidth	20 MHz
Noise of VLC channel	4.7×10^{-14} A^2
LTE channel bandwidth	10/20/30 MHz
LTE SINR	20 dB
VLC SINR threshold	10 dB

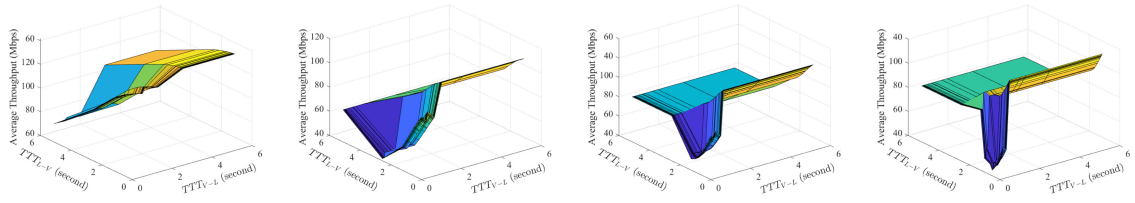


(a) LTE bandwidth = 10 MHz (b) LTE bandwidth = 20 MHz (c) LTE bandwidth = 30 MHz

Fig. 4. Evaluation of average throughput in terms of different values of TTT_{L-V} while the user crosses the VLC cell at 4 different speeds (i.e., 0.5, 1, 1.5 and 2 m/s). Three cases regarding the LTE bandwidth allocated to a single UE are evaluated: (a) 10 MHz; (b) 20 MHz; (c) 30 MHz.

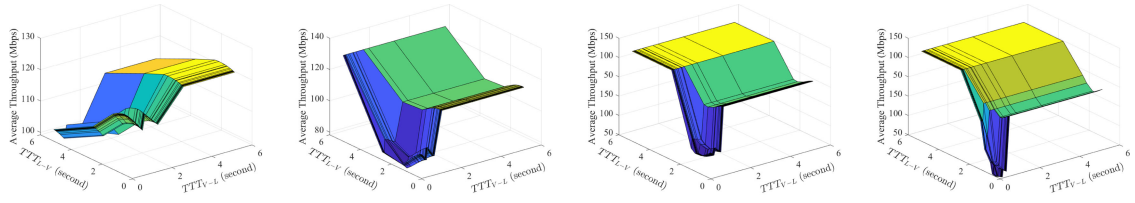
The 1 W here is the transmitted optical power rather than the consumed electrical power by the driver circuitry and the light source. According to the results in [25] (i.e., Fig. 10 in [25] shows the power consumption and Fig. 13 in [25] shows the illumination level), 1 W emitted optical power for each VLC AP in a grid structure with 3 meters as interval distance between adjacent VLC APs is sufficient to meeting the typical indoor illumination requirement (i.e., 300–500 lux). The effective sensing area, field-of-view (FOV) and optical-to-electrical (O/E) conversion efficiency factor are set to 10^{-4} m^2, 90° and 0.53 A/W, respectively. Given the VLC channel bandwidth as 20 MHz and noise power as 4.7×10^{-14} A^2, the SINR and the capacity of VLC channels are estimated by the path loss and Shannon capacity models in [25]. Considering the small coverage area of a single VLC AP, the variation of LTE SINR is relatively small while the user crosses the VLC cell due to the multipath effect. Therefore, the LTE SINR is set to 20 dB. To evaluate the impact of user movement pattern and LTE load on the average throughput while the user crosses the VLC cells, 4 different user movement speeds (i.e., 0.5, 1, 1.5 and 2 m/s) are considered under 3 different cases (i.e., 10, 20 and 30 MHz) of LTE bandwidth allocated to a single UE. The user movement trajectory is assumed to be a straight line crossing the centers of the VLC cells, which is the dominant indoor user movement pattern. The UE is initially connected to LTE and TTT_{L-V} will start counting once the SINR of VLC channel exceeds the VLC SINR threshold (i.e., 10 dB).

For the first system setting (i.e., one VLC AP), the value of TTT_{V-L} is always set to zero since the UE can only switch back to LTE after the SNR of the VLC channel drops below the threshold (i.e., 10 dB). The numerical results are shown in Fig. 4. It can be observed that the optimal TTT_{L-V} depends on both the user movement speed and LTE load. When the number of users associated with the LTE eNodeB is large, for example, during the working hours in an office building, the benefit obtained from handover into VLC is dominant. In Fig. 4(a), even if the user crosses the VLC cell at a relatively high speed (i.e., 2 m/s), handover into VLC immediately when the VLC SNR exceeds the threshold is still the better option than avoiding the handover by a large TTT_{L-V} such as 2.56 or 5.12 seconds. When the number of users associated with the LTE eNodeB is small, for example,



(a) User speed = 0.5 m/s (b) User speed = 1 m/s (c) User speed = 1.5 m/s (d) User speed = 2 m/s

Fig. 5. Evaluation of average throughput in terms of different values of TTT_{L-V} and TTT_{V-L} while the user is crossing two VLC cells at 4 different speeds: (a) 0.5 m/s; (b) 1 m/s; (c) 1.5 m/s; (d) 2 m/s. LTE bandwidth allocated to a single UE is set to 10 MHz.

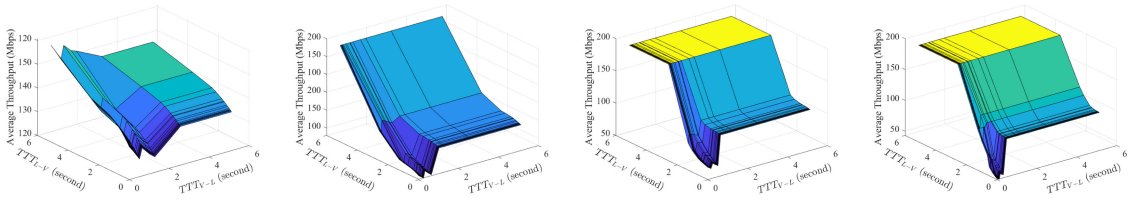


(a) User speed = 0.5 m/s (b) User speed = 1 m/s (c) User speed = 1.5 m/s (d) User speed = 2 m/s

Fig. 6. Evaluation of average throughput in terms of different values of TTT_{L-V} and TTT_{V-L} while the user is crossing two VLC cells at 4 different speeds: (a) 0.5 m/s; (b) 1 m/s; (c) 1.5 m/s; (d) 2 m/s. LTE bandwidth allocated to a single UE is set to 20 MHz.

during the early morning and late night, the benefit obtained from handover into VLC depends on the moving speed. In Fig. 4(b), if the user crosses the VLC cell slowly at a speed 0.5 m/s, the average throughput slightly increases along with the increase of TTT_{L-V} until $TTT_{L-V} = 1.28$ seconds. This is because before TTT_{L-V} reaches to 1.28 seconds, the handover process, during which the throughput is zero, is shifting to a time period which leads to the minimum throughput loss. Nevertheless, if the user moving speed exceeds 1 m/s, preventing the handover by a large TTT_{L-V} is always the best choice. Taking into account the carrier aggregation capability of LTE [26], the throughput gain of handover into VLC is further diluted. In Fig. 4(c), if the LTE bandwidth allocated to a single UE reaches 30 MHz, even if the mobile UE crosses the VLC cell at a low speed 0.5 m/s, the optimal TTT_{L-V} is a sufficiently large value to prevent the occurrence of handover. In summary, for the case of one VLC AP, the purpose of setting non-zero TTT_{L-V} for handover from LTE to VLC is mainly avoiding the handover when the user speed is high and LTE congestion level is low. The benefit of setting non-zero TTT_{L-V} in order to delay the handover is not remarkable.

For the second system setting (i.e., two VLC APs), we evaluate the average throughput in terms of different combinations of TTT_{L-V} and TTT_{V-L} with different user moving speed and LTE load. Both VLC APs transmit data on the same channel and cause inter-channel interference to each other. The numerical results are shown in Figs. 5, 6 and 7. All the results show that the difference of average throughput between the maximum and minimum is considerable which implies a fact that jointly optimizing the values of TTT_{L-V} and TTT_{V-L} is needed. In Fig. 5, the LTE bandwidth allocated to a single UE is relatively low (i.e., 10 MHz). Even if the user moving speed increases up to 2 m/s, the optimal pair of (TTT_{L-V}, TTT_{V-L}) is (0, 5.12), which indicates that the benefit acquired from associating with VLC APs is still dominant. However, as the user speed increases, the average throughput corresponding to small TTT_{L-V} and TTT_{V-L} values decreases significantly. This is because the proportion of the handover overhead during the user crossing the VLC cells becomes higher as the user speed increases. In Fig. 6, the LTE bandwidth allocated to a single UE is set to 20 MHz. If the user crosses the VLC cells at a low speed 0.5 m/s, the optimal pair



(a) User speed = 0.5 m/s (b) User speed = 1 m/s (c) User speed = 1.5 m/s (d) User speed = 2 m/s

Fig. 7. Evaluation of average throughput in terms of different values of TTT_{L-V} and TTT_{V-L} while the user is crossing two VLC cells at 4 different speeds: (a) 0.5 m/s; (b) 1 m/s; (c) 1.5 m/s; (d) 2 m/s. LTE bandwidth allocated to a single UE is set to 30 MHz.

of (TTT_{L-V}, TTT_{V-L}) is still (0, 5.12). Nevertheless, if the user speed exceeds 1 m/s, immediately handover into VLC (i.e., $TTT_{L-V} = 0$) once the condition is met always leads to a relatively low average throughput. It will be even worse if TTT_{V-L} is also equal to zero since the UE will be switched back and forth between LTE and VLC multiple times. In Fig. 7, the LTE bandwidth allocated to a single UE increases to 30 MHz. Under such condition, even if the user speed is set to 0.5 m/s, the optimal pair of (TTT_{L-V}, TTT_{V-L}) is (5.12, 0), which indicates that handover into VLC opportunistically becomes less beneficial than keep connecting to LTE. In summary, for the case of two VLC APs, when LTE load is high, typically TTT_{L-V} is set to zero and TTT_{V-L} is set to the maximum to keep UE connecting to the VLC. When LTE load is medium, the optimal TTT_{L-V} and TTT_{V-L} depends on the user speed. When LTE load is low, typically TTT_{L-V} is set to the maximum to avoid handover from LTE to VLC.

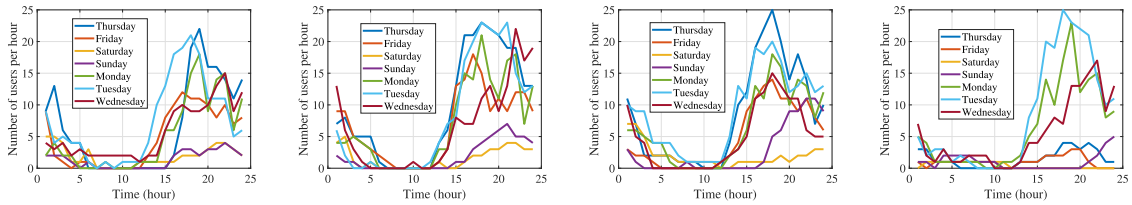
6. Real-Subject Evaluation

In this section, we evaluate the number of active users and the average moving speed per hour for everyday over a month based on real measurements [12], [13] conducted by the researchers from The University of New Mexico in a three-floor in-building environment. The dataset,[2] which is publicly available, includes the Raspberry Pi 3 (RPi) collection of Bluetooth Low Energy (BLE) advertisement packets gathered from 46 participants (i.e., faculties, staffs and students) carrying iBeacons and following their routines during a one-month period. Each iBeacon broadcasts packets every 1 second with omni-directional antenna propagation setting and transmission power of 0 dBm. Each received packet is reported to the centralized server with the beacon/user ID, the packets Received Signal Strength Indicator (RSSI), the corresponding timestamp and the ID of the RPi that received the advertisement packet.

To evaluate the number of active users (i.e., the participants who are carrying the iBeacons and staying in the tracking environment) per hour, we sort all the RSSI reports by the timestamps and extract the unique beacon/user IDs for each hour period. The results are shown in Fig. 8. The one-month period is divided into four weeks. For each week, the distribution of the number of active users over a day is plotted for the seven days. The distributions for almost all of the weekdays follow the same trend. It indicates the repetitiveness of the distribution of LTE spectral congestion level over a day, which is directly affected by the number of active users in the coverage of the LTE eNB.

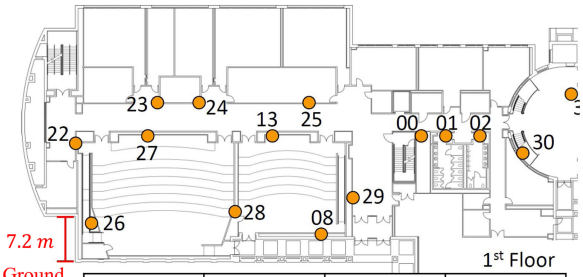
To evaluate the average moving speed of the participants, their real-time locations need to be estimated. Note that although it is presented in [12] that the dataset can be exploited for locating a user within a confidence interval, the 2D coordinates location tracking and mobility pattern estimation have not been studied by the contributors of the dataset. In order to estimate the locations of the participants, we first generate the 2D coordinates of all the RPis as shown in Fig. 9. The three floor plans including the locations of the RPi scanners (Fig. 3 in [12]) are converted

[2]https://github.com/dimisik/BLEBeacon-Dataset



(a) The first week (b) The second week (c) The third week (d) The fourth week

Fig. 8. The number of active users per hour in: (a) the first week; (b) the second week; (c) the third week; (d) the fourth week.



Coordinates	(1022,335)	(1085,335)	(1172,335)
RPi ID	00	01	02

Fig. 9. Coordinates of the RPis and the ground truth of distance.

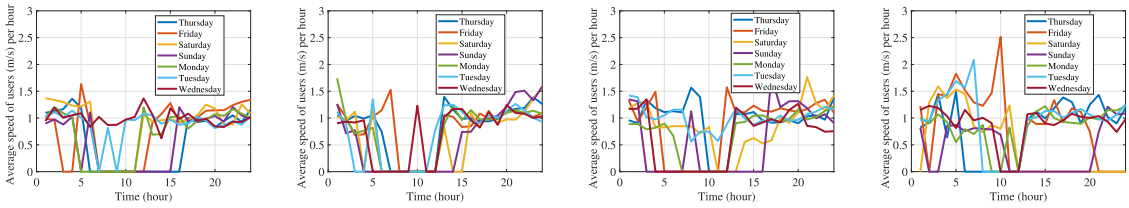
into images with grid of pixels. Based on the ground truth of the approximated scale (i.e., 7.2 m for 116 pixels), the unit of the coordinates of the RPis are changed from pixel to meter. For example, as shown in Fig. 9, the ground truth of the coordinate of the RPi 00 is $(1022, 335) \times \frac{7.2}{116} = (63.4, 20.8)$.

Due to the fact that reported packet includes RSSI and the corresponding timestamp, we apply RSSI and trilateration localization method to determine the location of the participant when there are more than two unique RPi IDs reported during the decision period. All the reported RSSI packets related to a specific beacon/user ID during the decision period are considered when determining the location of the user. The minimum decision period is empirically set to 10 seconds according to the experimental results presented in Fig. 5 in [13]. Having the ground truth of the coordinates of the RPis, the next step is to estimate the distance between a participant and a RPi according to the reported RSSI. Based on the curve fitting results of Fig. 3 and equation (1) in [13], the relationship between RSSI and the distance is estimated by $d = 10^{\frac{RSSI+C}{-10\gamma}} \times d_0$, where the path loss exponent $\gamma = 1.591$, the reference distance $d_0 = 5$ and the average value of RSSI at d_0 $C = 83.85$. Denote the coordinates of the user as (x_0, y_0) and the coordinates of the ith reported RPi during the decision period as (x_i, y_i), we have the distances between the user and the three reported RPis with the highest RSSI values as

$$d_i = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2},\ i = 1, 2, 3. \tag{4}$$

The location of the user can be determined by solving the above system equations. We turn the equation solving into an optimization process trying to minimize the square errors between the summations of the left and the right hand sides of (4),

$$\vec{s}_{\,opt} = \arg\min_{s} \sum_{i} f_i(\vec{s})^2, \tag{5}$$



(a) The first week (b) The second week (c) The third week (d) The fourth week

Fig. 10. The average user moving speed (>0.5 m/s) per hour in: (a) the first week; (b) the second week; (c) the third week; (d) the fourth week.

where $\overrightarrow{s} = (x_0, y_0)$ and

$$f_i(\overrightarrow{s}) = d_i - \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2},\ i = 1, 2, 3. \tag{6}$$

Sometimes the number of unique RPi IDs during the decision period may not be larger than or equal to three. For the case when there is only one unique RPi ID, the location of the user is set to the same coordinates of the corresponding RPi. For the case when there are two unique RPi ID, the location of the user is set to the middle point of the two corresponding RPis. In addition to the problem of the number of unique RPi IDs, the floor bleeding also creates troubles to the location tracking. The floor bleeding is referred to as the events where a scanner from another floor picked up the advertisement packet. To handle this problem, during each decision period, we use the floor that has the most number of reported RPi IDs as the current floor where the user is located at.

After the location of the user is determined for each decision period, the location is associated with the timestamp at the end of each decision period. Given that the locations at time t_1 and t_2 are (x_1, y_1) and (x_2, y_2), respectively, the speed of user at t_1 is calculated by $v = \frac{\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}}{t_2-t_1}$. Note that the value of $t_2 - t_1$ may be much longer than 10 seconds since the participants were not continuously staying in the building. However, under such a condition, the speed estimation will be close to zero, which does not negatively affect the average speed evaluation given that we set a minimum speed threshold (i.e., 0.5 m/s) when we average the results. Also note that there exist some cases where the user is located at different floors at t_1 and t_2. For such cases, we add 20 meters and 40 meters to the traveling distance when the floor differences are 1 and 2, respectively. The generated speed of users are averaged per hour for the one-month experimental period. As shown in Fig. 10, the distributions of the average speed per hour for almost everyday follow the same trend especially during the working hours. The results validate the repetitiveness of indoor user moving speed over a day. The mobility patterns acquired from the real data set let us generate very practical estimation of the outcome of the learning algorithm given the fact that the channel characteristics of LTE and VLC have been well studied.

7. Simulation Results

In this section, simulations[3] are conducted in MATLAB R2018 to evaluate the impact of different values of exploration and exploitation trade-off factor ϵ, TTT space and number of time slots t_s on the optimal throughput and convergence speed. A converged Q-table running in an online manner is also evaluated under different ϵ control schemes and different TTT space. The converged throughput of our proposed Q-learning based algorithm is compared to those of fixed TTT scheme and optimal performance under different t_s.

The simulation and Q-learning parameters are summarized in Table 2. One LTE eNodeB and two VLC APs are considered. Even though the number of candidate values of TTT_{L-V} and TTT_{V-L} will

[3]https://github.com/Sihua-Shao/Handover

TABLE 2
Simulation Parameters

TTT range	$[0.0\ 5.12]$ sec
Number of time slots	$12/24/48/96$
ϵ_{init}	$0.2/0.4/0.8/1$
ϵ control factor a	500
ϵ control factor b	1
Learning rate α	1
Discount factor γ	0.9
Max number of episodes	10000
Initial state s_{init}	$(5.12\ sec, 0\ sec)$
Mean of user movement speed	$1.5 - e^{-\frac{(t-ts/2)^2}{2(t_s/4)^2}}$ m/s
LTE bandwidth	$20 \times [1.1 - e^{-\frac{(t-ts/2)^2}{2(t_s/4)^2}}]$ MHz

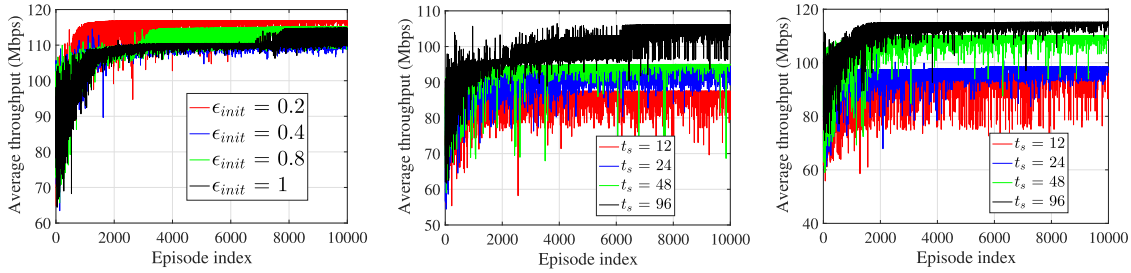


Fig. 11. (a) Evaluation of convergence speed in terms of different initial ϵ values. (b) Evaluation of optimal throughput and convergence speed in terms of different number of time slots t_s when the TTT space is large. (c) Evaluation of optimal throughput and convergence speed in terms of different number of time slots t_s when the TTT space is small.

be changed, they are still in the range of [0 5.12]. 4 different numbers of time slots (i.e., 12, 24, 48 and 96) are evaluated. For example, if $t_s = 12$, the length of each time slot is 2 hours. Four different initial ϵ values are considered. $\epsilon_{init} = 1$ means the action will always be selected randomly at the beginning of the training process. The initial state s_{init} is set to (5.12, 0) since in the early morning the LTE load is typically low and user speed is high. The mean of user speed and LTE bandwidth allocated to a single UE are Gaussian functions of t, where the user speed is low and LTE load is high in the midday.

In Fig. 11(a), the number of time slots $t_s = 96$, 16 candidate values of both TTT_{L-V} and TTT_{V-L} are considered, and 4 different initial ϵ values are evaluated. If $\epsilon_{init} = 1$, the value of ϵ will decrease to around 0.05 at episode 10000. It can be seen that allowing high exploration capability may not lead to efficient convergence. Therefore, ϵ_{init} is set to 0.2 for the simulations evaluating different t_s and TTT space.

In Fig. 11(b), $\epsilon_{init} = 0.2$, 16 candidate values of both TTT_{L-V} and TTT_{V-L} are considered, and 4 different t_s are evaluated. When $t_s = 12$, the Q-table converges fast but the optimal throughput is very low. This is because the values of TTT_{L-V} and TTT_{V-L} have to be fixed for 2 hours which is sufficiently long for the user movement pattern and LTE load to change significantly. When $t_s = 96$, even though the optimal throughput is improved when compared to that of the smaller values t_s, the convergence speed is notably deteriorated.

According to the results in the case studies (Section 5), the optimal TTT_{L-V} and TTT_{V-L} generally is located at extreme values (i.e., 0 or 5.12). Therefore, in Fig. 11(c), all the settings are the same as those in Fig. 11(c) except the TTT space is reduced from 16 candidate values (i.e., 0, 0.04, 0.064, 0.08, 0.1, 0.128, 0.16, 0.256, 0.32, 0.48, 0.512, 0.64, 1.024, 1.28, 2.56 and 5.12 seconds) to 9

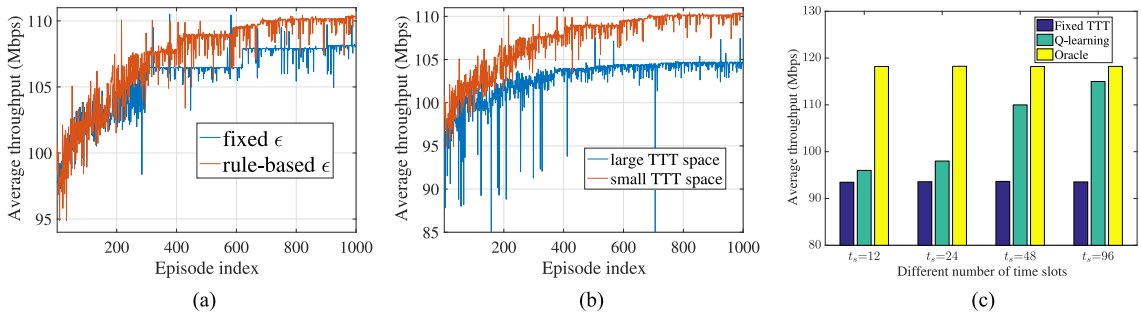


Fig. 12. (a) Online evaluation of convergence speed in terms of different ϵ control schemes. (b) Online evaluation of convergence speed in terms of different TTT space (c) Optimal throughput comparison among fixed TTT scheme, Q-learning based algorithm and oracle.

candidate values (i.e., 0.0 0.04 0.08 0.16 0.32 0.64 1.28 2.56 and 5.12 seconds). As the outcome of reducing the TTT space, both the convergence speed and the optimal throughput are improved especially for large t_s. The reason is that it takes less number of time slots to change the values of TTT_{L-V} and TTT_{V-L} from one extreme to the other extreme.

Since our final goal is to design a self-optimizing algorithm running on the CC while operating the LTE-VLC HetNets, we first train the Q-table through offline emulator which emulates the estimated distribution of user movement pattern and LTE load and then apply the converged Q-table to the Q-learning agent running in an online manner. In Fig. 12(a), we change the deviation of user speed and LTE bandwidth from $t_s/4$ to $t_s/2$ and compare the performance of two schemes: i) fixed ϵ - ϵ is fixed at 0.05 to allow certain exploration capability; ii) rule-based ϵ - upon the detection of non-negligible throughput drop, ϵ is initialized as 0.2 and decreases along with the number of episodes. The performance of two schemes are similar and the rule-based ϵ scheme slightly outperforms the fixed ϵ scheme.

In Fig. 12(b), the deviation of user speed and LTE bandwidth is also changed from $t_s/4$ to $t_s/2$, rule-based ϵ scheme is adopted and the performance of large TTT space (i.e., 16 candidate values) and small TTT space (i.e., 9 candidate values) are compared. By reducing the size of TTT space, the online performance of the algorithm is also improved in terms of convergence speed and optimal throughput. In addition, during the online Q-table training process, the average throughput of small TTT space will not decrease to a very low value (e.g., below 90 Mbps) as large TTT space will do.

In Fig. 12(c), we compare the optimal performance of three schemes: i) fixed TTT - the values of TTT_{L-V} and TTT_{V-L} are both fixed at zero; ii) Q-learning - the values of TTT_{L-V} and TTT_{V-L} are optimized by our proposed Q-learning based algorithm for each time slot; iii) Oracle - the values of TTT_{L-V} and TTT_{V-L} are optimal in each time slot. When $t_s = 12$, the optimal performance of Q-learning based algorithm is close to that of fixed TTT scheme. However, as t_s increases, the optimal performance of Q-learning based algorithm is approaching to that of oracle. When $t_s = 96$, Q-learning based algorithm improves the average throughput by 25% when compared to the fixed TTT scheme.

8. Conclusion

In order to resolve the limitations of the quasi-static network model and the immediate handover policy from RF to VLC, in this paper, we propose a flexible and holistic framework to enable the self-optimization of handover parameters in LTE-VLC HetNets. A Q-learning based algorithm is designed to optimize a sequence of (TTT_{L-V}, TTT_{V-L}) for the purpose of maximizing the average throughput especially for the mobile UEs. Case studies are performed to evaluate the gain acquired from optimizing TTT_{L-V} and TTT_{V-L}. Simulation results indicate that the Q-learning based algorithm

improves the average throughput by 25% when compared to the fixed TTT scheme. As the resolution of controlling TTT values increases, the optimal throughput becomes comparable to the oracle performance while convergence speed is compromised.

References

[1] Huawei. 2016. [Online]. Available: https://www.huawei.com/minisite/hwmbbf16/insights/5G-Nework-Architecture-Whitepaper-en.pdf. Accessed: Sep. 1, 2018.

[2] M. Ayyash *et al.*, "Coexistence of WiFi and LiFi toward 5G: Concepts, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 64–71, Feb. 2016.

[3] S. Shao *et al.*, "Design and analysis of a visible-light-communication enhanced WiFi system," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 7, no. 10, pp. 960–973, Oct. 2015.

[4] Z. Du, C. Wang, Y. Sun, and G. Wu, "Context-aware indoor VLC/RF heterogeneous network selection: Reinforcement learning with knowledge transfer," *IEEE Access*, vol. 6, pp. 33275–33284, 2018.

[5] Y. Wang, X. Wu, and H. Haas, "Fuzzy logic based dynamic handover scheme for indoor Li-Fi and RF hybrid network," in *Proc. IEEE Int. Conf. Commun.*, 2016, pp. 1–6.

[6] L. Li, Y. Zhang, B. Fan, and H. Tian, "Mobility-aware load balancing scheme in hybrid VLC-LTE networks," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2276–2279, Nov. 2016.

[7] J. Zhang, X. Zhang, and G. Wu, "Dancing with light: Predictive in-frame rate selection for visible light networks," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 2434–2442.

[8] R. Liu and C. Zhang, "Dynamic dwell timer for vertical handover in VLC-WLAN heterogeneous networks," in *Proc. IEEE 13th Int. Wireless Commun. Mobile Comput. Conf.*, 2017, pp. 1256–1260.

[9] S. Liang, Y. Zhang, B. Fan, and H. Tian, "Multi-attribute vertical handover decision-making algorithm in a hybrid VLC-Femto system," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1521–1524, Jul. 2017.

[10] X. Bao, J. Dai, and X. Zhu, "Visible light communications heterogeneous network (VLC-HetNet): New model and protocols for mobile scenario," *Wireless Netw.*, vol. 23, no. 1, pp. 299–309, 2017.

[11] D. Lopez-Perez, I. Guvenc, and X. Chu, "Mobility management challenges in 3GPP heterogeneous networks," *IEEE Commun. Mag.*, vol. 50, no. 12, pp. 70–78, Dec. 2012.

[12] M. Inaya, M. Meli, D. Sikeridis, and M. Devetsikiotis, "A real-subject evaluation trial for location-aware smart buildings," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2017, pp. 301–306.

[13] D. Sikeridis, M. Devetsikiotis, and I. Papapanagiotou, "Occupant tracking in smart facilities: An experimental study," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2017, pp. 818–822.

[14] F. Wang, Z. Wang, C. Qian, L. Dai, and Z. Yang, "Efficient vertical handover scheme for heterogeneous VLC-RF systems," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 7, no. 12, pp. 1172–1180, Dec. 2015.

[15] M. S. Saud and M. Katz, "Implementation of a hybrid optical-RF wireless network with fast network handover," in *Proc. VDE 23th Eur. Wireless Conf.*, 2017, pp. 1–6.

[16] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmwave mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2069–2084, Sep. 2017.

[17] R. Alkhansa, H. Artail, and D. M. Gutierrez-Estevez, "LTE-WiFi carrier aggregation for future 5G systems: A feasibility study and research challenges," *Elsevier Procedia Comput. Sci.*, vol. 34, pp. 133–140, 2014.

[18] F. Guidolin, I. Pappalardo, A. Zanella, and M. Zorzi, "Context-aware handover policies in HetNets," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1895–1906, Mar. 2016.

[19] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100–107, Feb. 2004.

[20] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[21] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[22] J. Wu, J. Liu, Z. Huang, and S. Zheng, "Dynamic fuzzy Q-learning for handover parameters optimization in 5G multi-tier networks," in *Proc. IEEE Int. Conf. Wireless Commun. Signal Process.*, 2015, pp. 1–5.

[23] *Evolved Universal Terrestrial Radio Access (E-UTRA)*, ETSI, TS, 136 331 V13. 0.0 LTE, Jan. 2016.

[24] I. Abdalla, M. Rahaim, and T. Little, "Impact of receiver FOV and orientation on dense optical networks," in *Proc. IEEE Global Commun. Conf.*, 2018, pp. 1–6.

[25] S. Shao, A. Khreishah, and I. Khalil, "Joint link scheduling and brightness control for greening VLC-based indoor access networks," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 8, no. 3, pp. 148–161, Mar. 2016.