Reinforcement Learning Approach for Hybrid WiFi-VLC Networks

Abdulmajeed M. Alenezi
School of Electrical and Electronic Engineering
The University of Manchester
Manchester, UK
Abdulmajeed.Alenezi@manchester.ac.uk

Khairi A. Hamdi
School of Electrical and Electronic Engineering
The University of Manchester
Manchester, UK
K.Hamdi@manchester.ac.uk

***Abstract*—The number of mobile devices in indoor environment has dramatically increased and the capacity of conventional RF wireless networks may not be enough to support the indoor traffic demand. Recently, Visible Light Communication (VLC) systems have emerged as a complementary unlicensed media. In this paper, we proposed a hybrid WiFi-VLC system where multiple VLC access points (AP) coexist with a WiFi AP. A number of indoor users share the hybrid WiFi-VLC system. All users employ WiFi for uplink whereas one access point (WiFi or VLC) is assigned for each user to maximize the overall capacity of the network. We propose a new reinforcement learning algorithm which can be implemented at the WiFi AP and results in the selection of an access point such that the total throughput is maximized. Numerical simulation results show that the proposed method improves the total system throughput significantly. Furthermore, the throughput achieved by the worst user in the proposed Q-Learning algorithm becomes higher than what would be received by the average user who used the conventional hybrid systems based on best connection.**

***Index Terms*—WiFi, VLC, Reinforcement Learning, Hybrid system, Centralized Q-Learning**

I. Introduction

Future wireless networks are expected to maintain the quality of service for all users despite the dramatic increase in mobile devices, especially in indoor environment [1]. Maintaining the required high data rate with low delay for a large number of users might not be applicable in the current systems due to the limitation in the radio frequency (RF) spectrum. One possible way to improve the indoor wireless network is by using hybrid system with multiple networks. To support multiple networks, multi-homing capability has been developed which allows the user to receive data from multiple networks [2].

Selecting the best complementary network is crucial as it might significantly increase the hybrid system's complexity, which results in more complicated schemes. For example, in hybrid LTE-WiFi, both networks operate at the same frequency, which increases the co-channel interference. As both networks share the same spectrum, the use of a hybrid system might not significantly improve the overall performance. Most of the research on hybrid LTE-WiFi focus on improving the energy efficiency which is out of the scope in this paper [3].

Visible light communication (VLC) recently brought a great potential as a complementary network to the WiFi due to many factors such as low energy consumption, unlicensed band, and security [4]. As the VLC can be directed, it makes it suitable for achieving a high data rate on a small area of coverage. However, VLC is mainly implemented in the downlink, and it needs a reliable uplink connection such as WiFi or infrared since it is not practical to be used in the uplink [5]. Combining WiFi and VLC would benefit from both networks' advantages and overcome the limitations of both networks.

There have been several studies that investigate the implementation of heterogeneous RF and VLC network [5]- [7]. In [5], the authors proposed a heterogeneous system in which the WiFi is used in the uplink while the VLC is used in the downlink. In this case, the hybrid system improved the overall performance but did not reach the full potential of using WiFi-VLC in the downlink. In [6], the authors investigate the handover mechanism in hybrid RF-VLC while [7] investigates the energy efficiency of the hybrid system.

Recently, several studies suggest the use of reinforcement learning in hybrid networks [8] - [10]. The authors in [8] apply reinforcement learning on hybrid LTE, WLAN, and VLC for network selection. The author take into consideration the traffic type and the possibility of having learning records to improve the Q-learning algorithm. In [9], the authors proposed a new reinforcement learning algorithm for energy efficient resource management. In [10], the authors use multi-agent reinforcement learning to develop online power allocation that improves the user's QoS.

In this paper, we propose a new centralized Q-learning algorithm on the WiFi AP that improves the total system performance. Our contribution is categorized into two main points.

- The resource allocation problem in a hybrid WiFi-VLC system is solved using centralized Q-learning. The proposed algorithm offload users from one AP to another to improve the overall QoS.
- A new reward function that takes into consideration the user location to minimize the handover.

The rest of the paper is organized as follows. Section II describes the WiFi, VLC and the hybrid model. Problem formulation and the proposed reinforcement learning approach is presented in section III and IV, respectively. Section V presents simulation results and finally, section VI concludes the paper.

978-1-7281-5207-3/20/$31.00 ©2020 IEEE

II. System Model

We consider an indoor heterogeneous wireless access environment, which consists of one WiFi access point, M VLC access points, and K users. All users are equipped with multi-homing capability and can only connect to one AP. The uplink is served by the WiFi, while the downlink can be served by either VLC or WiFi. As shown in Fig. 1, some users might connect to the WiFi even though they are located under the VLC AP to maximize the total system performance. VLC system significantly different from the RF systems in terms of operating frequency and modulation/demodulation techniques which make it suitable for a hybrid system with WiFi as both operates at different frequencies.

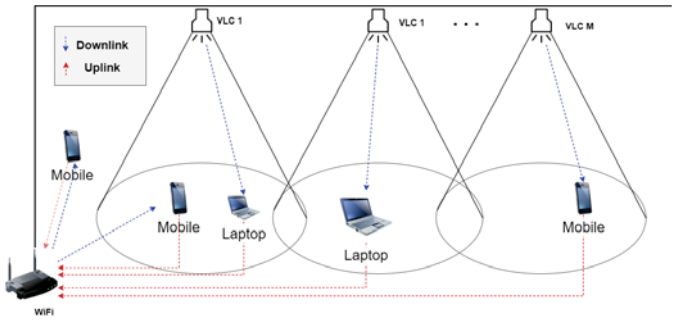

Fig. 1. System architecture of hybrid VLC-WiFi network

A. WiFi Model

Due to the implementation of CSMA/CA schedule scheme in 802.11, each user occupies the total bandwidth for a time interval t. Thus, the user throughput can be calculated by averaging in a period of time T [11].

$$C^k_{\mathrm{WiFi}} = s_k[B\log_2(1+\mathrm{SNR_k})] \tag{1}$$

where $s_k \in [0,1]$, which corresponds to the time interval user k occupied the channel over the total time ($\frac{t_k}{T}$). Note that $\sum_{n=1}^{N}\frac{t_n}{T}=1$. As there is only one WiFi AP and both VLC and WiFi operates at different frequency, there is no co-channel interference and the Signal to Noise Ratio (SNR) for user k is given by:

$$\mathrm{SNR^k} = \frac{P_t \cdot h_{k,l}}{N_0} \tag{2}$$

where P_t is the transmitted power, $h_{k,l}$ is the channel gain between the WiFi AP and user k, and N_0 is the PSD of noise at the receiver.

B. VLC Model

The VLC system consists of multiple LED lamps operate as VLC access points. Each lamp is considered as a single AP and contains multiple Q LEDs with different orientations. As shown in Fig 2, the line of sight (LOS) channel gain between the n^{th} user's photodetector and the q^{th} LED is given by [12]

$$h_{qn} = \frac{R_e A_r \cos(\phi)}{2\pi d^2_{qn}}(m+1)\cos^m(\psi) \tag{3}$$

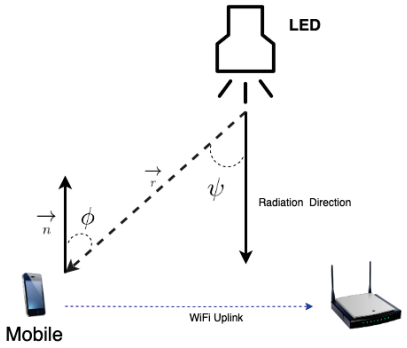

Fig. 2. VLC LOS Downlink

m is the order of Lambertian mode for the light source, which is related to the LED's semiangle $\Phi_{\frac{1}{2}}$ by $m = \frac{\ln 2}{\ln(cos(\Phi_{\frac{1}{2}}))}$. Table (I) illustrates the rest of the notations in (3).

TABLE I
Notation in Equation 3

Head	Head
$\overrightarrow{r_{qk}}$	Unit vector pointing toward user k from LED q
$\overrightarrow{n_k}$	Normal unit vector for the k_{th} user
$\overrightarrow{l_k}$	Radiation unit vector for the k_{th} user
ϕ	The angle between $\overrightarrow{r_{qk}}$ and $\overrightarrow{n_k}$ vectors
ψ	The angle between $\overrightarrow{r_{qk}}$ and $\overrightarrow{l_k}$ vectors
A	Area of the photodetector
d_{qk}	Distance between transmitter q and user k
m	Lambertian mode of the light source
R_e	Responsivity of the photodetector

The average received optical signal is the sum of all received power from all the LEDs in the VLC AP. Similar to [13], the VLC channel characteristics have been simplified.

$$P^k_{r_{\mathrm{VLC}}} = \sum_{q=1}^{Q} h_{kq} P_{t_{\mathrm{VLC}}} \tag{4}$$

where $P_{t_{\mathrm{VLC}}}$ is the transmitted power for the VLC AP. To calculate the signal to noise ratio (SNR), we need to convert the optical power to electrical power by multiplying the optical power by the responsivity of the photodetector at the receiver.

$$\mathrm{SNR^k_{VLC}} = \frac{(R_e P^k_{r_{\mathrm{VLC}}})^2}{N_{0_{\mathrm{VLC}}} B_{\mathrm{VLC}}} \tag{5}$$

Since VLC uses intensity modulation and direct detection for optical signals, half of the subcarriers are used after modulation as only the real valued signals can be transmitted. When VLC AP support multiple users, TDMA with RR scheduling is used to support the assigned users [7]. Thus, the achievable rate for user k when connected to VLC AP l is given by

$$C^k_{\mathrm{VLC}_l} = \frac{B_{\mathrm{VLC}}}{2U_l}\log_2(1+\mathrm{SNR}_k) \tag{6}$$

where U_l is the total number of users assigned to the same AP.

III. Problem Formulation

In a hybrid WiFi/VLC system, the WiFi covers a large area and it is assumed that all users can connect to. However, due to the fairness in WiFi, users who are located far from the AP will take more time to transmit compare to the users who are close to the AP. By offloading WiFi users to the VLC APs, the system performance significantly improve as the users are distributed over multiple APs. Note that each user can connect to only one AP at time t. The total throughput for WiFi is given by

$$C_{\mathrm{WiFi}} = \sum_{k=1}^{K} s_k(B\log_2(1+\mathrm{SNR}_k)) \tag{7}$$

Similarly, the total throughput for one VLC AP is given by

$$C_{\mathrm{VLC}_l} = \sum_{k=1}^{K} \frac{B}{2U_l}\log_2(1+\mathrm{SNR}_k) \tag{8}$$

The total system throughput can be calculated by adding (7) and (8)

$$C_{total} = C_{\mathrm{WiFi}} + \sum_{m=1}^{M} C_{\mathrm{VLC}}^{M} \tag{9}$$

where M is the total number of VLC APs. Let L be the total number of available APs in the hybrid system. Since each user can connect to only one AP at time t, (9) needs to follow the constraint $\delta_k^l = 1$ for only AP l and 0 for the other APs, which means that user k is connected to AP l. The goal is to maximize the system throughput by reassigning users to each AP so that we can achieve higher total throughput. To do that, (7) and (8) can be rewritten as

$$\begin{aligned} C_{\mathrm{WiFi}} &= \sum_{k=1}^{K} \delta_k^{\mathrm{WiFi}} s_k(B\log_2(1+\mathrm{SNR}_k)) \\ C_{\mathrm{VLC}_l} &= \sum_{k=1}^{K} \delta_k^l \frac{B}{2U_l}\log_2(1+\mathrm{SNR}_k) \end{aligned} \tag{10}$$

where U_l is the total number of connected users to AP l. The maximum total throughput is given by

$$\max_{\delta_k^{\mathrm{WiFi}},\delta_k^{\mathrm{VLC}_1},\ldots,\delta_k^{\mathrm{VLC}_L}} \left(C_{\mathrm{WiFi}} + C_{\mathrm{VLC}}^1 + ... + C_{\mathrm{VLC}}^L\right) \tag{11}$$

Solving (11) using exhaustive research is not practical as it can't support dense users environment. One approach that can be used to solve the optimization problem is by using reinforcement learning.

IV. Centralized Reinforcement Learning Approach

Since all users use the WiFi AP as uplink, centralized reinforcement learning can be applied at the WiFi AP using a controller to offload users from one AP to another in the downlink. In a heterogeneous network, the Q-learning parameters can be defined as:

- **Agent**: The WiFi AP acts as an agent as all users use it for uplink. The agent uses ϵ-greedy policy for exploration by choosing an action with a probability of $1-\epsilon$, and acting randomly with a probability of ϵ.
- **Actions**: For each user, the controller selects one action from a set of actions $A = (a_1, a_2, ..., a_L)$. The number of actions is the same as the total number of APs in the system and all actions have the same probability. Each action consists of a vector indicating the user should connect to which AP as shown in table II. Simply $a_{(1)}$ means that the user is connected to only WiFi while selecting action $a_{(L)}$ for user k allows the user to connect to only $\mathrm{VLC}^{(L)}$.

TABLE II
Set of actions

	WiFi	**VLC$^{(1)}$**	**VLC$^{(2)}$**	. . .	**VLC$^{(L)}$**
a_1	1	0	0	. . .	0
a_2	0	1	0	. . .	0
a_3	0	0	1	. . .	0
. . .	. . .	. . .	. . .	. . .	. . .
a_L	0	0	0		1

- **Reward function**: Defining the reward function significantly affects the system performance as it can be designed to satisfy specific goal. We propose a new reward function that can maximize the total system throughput and can be implemented for a centralized Q-learning approach. The reward function for user k selecting action a_k at time step t can be defined as

$$\begin{aligned} R_k = a_k(1)C_{WiFi} + a_k(2)g_{(1)}C_{VLC_1} + ... \\ + a_k(L)g_{(L)}C_{VLC_L} \end{aligned} \tag{12}$$

where $g_l = (\frac{v}{d^k_{\mathrm{VLC}_l}})$, $d^k_{\mathrm{VLC}_l}$ is the distance between VLC_l and user k, and v is a reference distance. g_l is used to imply higher reward value for assigning user k to VLC_l when the user is located close to the same AP. Once the distance is more than v meters, the reward value for assigning the user to VLC_l is significantly reduced because it is not reliable to connect to far VLC AP.

Once the reward value for each connected user is calculated, we apply the sum of the reward values in the Q update equation below

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_a * Q(s^{'},a) - Q(s,a)] \tag{13}$$

where γ is the discount rate, α is the learning rate, and r is the sum of reward values for each connected user.

$$r = \sum_{k=1}^{K} R_k \tag{14}$$

To obtain the optimal Q-value, the agent receives a reward r from selecting action a, which is affected by a discount factor γ for performing the policy. The Q-learning algorithm is shown in Algorithm (1) [14]. The main advantage of this approach is selecting the best network for each user is not based on only the individual user's preference but also on the overall performance. For example, let's assume only two users request transmission, and both are located under the same VLC AP. The algorithm might allow one user to transmit using VLC and the other one on WiFi, which benefits both users instead of sharing the same resources.

Algorithm 1 Q-Learning algorithm

initialize $Q(x_t, a_t)$ arbitrarily
for all episodes **do**
Initialize x_t

for all steps of episode **do**
Choose a_t for all users from set of actions
Take action a_t, observe R_t, x_{t+1}
$Q(x_t, a_t) \leftarrow (1 - \alpha)Q(x_t, a_t) + \alpha \max_a(R_t + \gamma Q(x_{t+1}, a))$
$x_t \leftarrow x_{t+1}$
end for
end for

V. Simulation Results

An indoor environment is simulated in a 5 m $\times$ 5 m $\times$ 3 m room size using one WiFi AP and two VLC APs. The user achieves an instantaneous rate based on his location and the fading parameters. Each VLC AP covers a small area of the room while the WiFi covers the entire room. The users are uniformly distributed in the room, and all the results are averaged over 40 runs. The parameters used in the VLC are summarized in table III. The WiFi is assumed to operate at 2.4 GHz, and the channel gain is assumed to depend only on the path loss. The path loss for the WiFi is given by [15]

$$L(d) = (\frac{4\pi d_1}{\lambda})^2 (\frac{d}{d_1})^n \tag{15}$$

where d_1 is a reference distance, d is the distance between the WiFi AP and the user, λ is the wavelength, and n is the path loss exponent. The parameters used in WiFi are summarized in table III.

In the proposed Q-learning method, we set the maximum number of iterations to 60000. The agent is the WiFi AP, and it uses ϵ-greedy policy with $\epsilon = 0.1$. The learning rate and the discount factor are set to be 0.5 and 0.9, respectively. The total number of actions in this scenario is three, as the user

TABLE III
Parameters

VLC Parameters	**Value**
P_t^{VLC}	25 W
A_{pd}	1 cm^2
Semi-angle	60°
Responsivity	0.5 A/W
B_{VLC}	20 MHz
$N_{0_{\mathrm{VLC}}}$	10^{-18} W/Hz
WiFi Parameters	**Value**
Path loss exponent	3
d_1	8 m
B_{WiFi}	20 MHz
$N_{0_{\mathrm{WiFi}}}$	-174 dBm/Hz

can connect to one of the three APs. For all connected users, the algorithm runs through all the iterations and updates the Q-value using the reward function in (12). Note that the Q-value in the centralized Q-learning at each iteration depends on the reward value for all connected users. To have a fair comparison, the proposed Q-learning algorithm is compared to a non reinforcement learning approach where the user is connected to the AP based on the best signal strength. In the rest of the simulation, we call the proposed method 'Proposed Q-Learning' while the conventional approach is called 'Algorithm 2: Best connection'. The WiFi stand alone performance is also simulated.

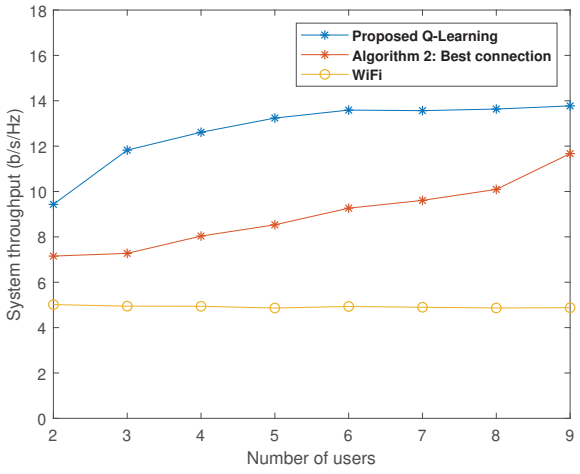

Fig. 3. Total system throughput comparison for different number of connected users

Fig. 3 shows the total system throughput for a various number of connected users. It is clear that using a hybrid system improves the network performance as both hybrid systems outperformed the WiFi stand alone. Compare to the WiFi stand alone, the proposed algorithm improved the system throughput by approximately 182% when the number of connected users is nine, while algorithm 2 improves the total throughput by approximately 139%. We can notice that connecting to the network with the highest signal strength is

not always the best case as offloading users using the proposed algorithm improves the total throughput significantly.

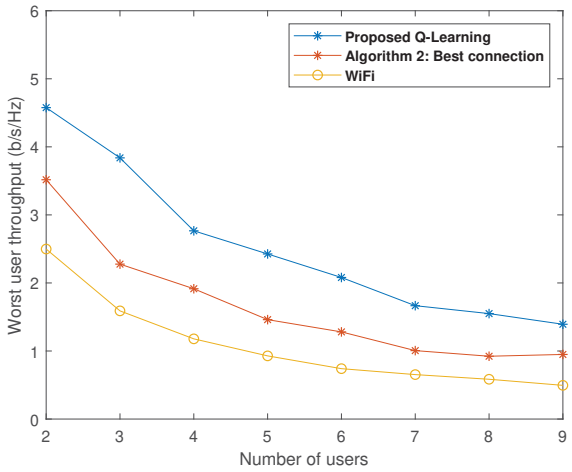

Fig. 4. Worst user throughput for different number of connected users

Another factor we need to take into consideration when testing the hybrid systems is the fairness among all connected users. In some cases, depending on the user's preference might affect the other users' performance. As shown in Fig. 4, connecting to the network with the best connection is not always the best option for maximizing the individual user throughput. The proposed algorithm shows an improvement in terms of maximizing the worst user's throughput.

Taking into consideration both the average system throughput and the worst user throughput, the proposed algorithm shows significant improvement in all cases as shown in Fig. 5. The worst user throughput in the proposed algorithm is better than what an average user would achieve in algorithm 2. As the number of users increase, algorithm 2 fails to maintain the fairness as the gap between the worst user throughput and the average user throughput increased. The proposed algorithm maintains the same level in all cases.

VI. Conclusion

Using hybrid WiFi-VLC system has a great potential to improve the indoor RF wireless networks as it overcomes the limitations of each network, such as low coverage and limited spectrum. In this paper, we analyzed the performance of the hybrid WiFi-VLC system using reinforcement learning. The algorithm is applied at the WiFi AP and improved the network selection by offloading users to one of the available APs. Numerical simulation results show a significant improvement in the total system throughput. The proposed algorithm also improves the fairness among all connected users in comparison to the conventional algorithm.

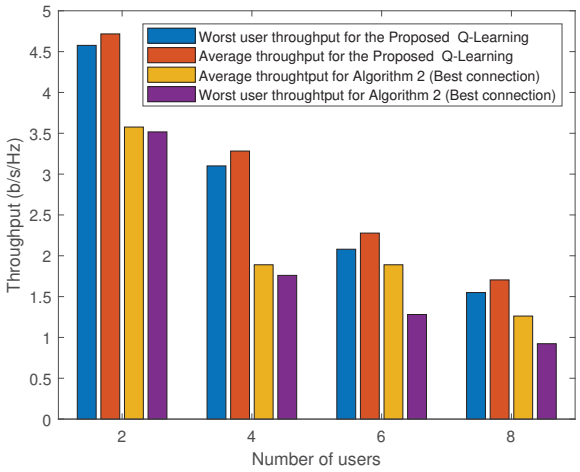

Fig. 5. Comparison of systems performance versus different number of connected users

References

[1] Ayyash, M., Elgala, H., Khreishah, A., Jungnickel, V., Little, T., Shao, S., Rahaim, M., Schulz, D., Hilt, J. and Freund, R., 2016. Coexistence of WiFi and LiFi toward 5G: concepts, opportunities, and challenges. IEEE Communications Magazine, 54(2), pp.64-71.

[2] Ismail, M. and Zhuang, W., 2014. Green radio communications in a heterogeneous wireless medium. IEEE Wireless Communications, 21(3), pp.128-135.

[3] Zhou, F., Feng, L., Yu, P. and Li, W., 2015, March. Energy-efficiency driven load balancing strategy in LTE-WiFi interworking heterogeneous networks. In 2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW) (pp. 276-281). IEEE. Vancouver

[4] Ghassemlooy, Z., Alves, L.N., Zvanovec, S. and Khalighi, M.A. eds., 2017. Visible light communications: theory and applications. CRC press.

[5] Rahaim, M.B., Vegni, A.M. and Little, T.D., 2011, December. A hybrid radio frequency and broadcast visible light communication system. In 2011 IEEE GLOBECOM Workshops (GC Wkshps) (pp. 792-796). IEEE.

[6] Shao, S., Khreishah, A., Ayyash, M., Rahaim, M.B., Elgala, H., Jungnickel, V., Schulz, D., Little, T.D., Hilt, J. and Freund, R., 2015. Design and analysis of a visible-light-communication enhanced WiFi system. Journal of Optical Communications and Networking, 7(10), pp.960-973.

[7] Basnayaka, D.A. and Haas, H., 2015, May. Hybrid RF and VLC systems: Improving user data rate performance of VLC systems. In 2015 IEEE 81st Vehicular Technology Conference (VTC Spring) (pp. 1-5). IEEE.

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

[9] Yang, H., Alphones, A., Zhong, W.D., Chen, C. and Xie, X., 2019. Learning-Based Energy-Efficient Resource Management by Heterogeneous RF/VLC for Ultra-Reliable Low-Latency Industrial IoT Networks. IEEE Transactions on Industrial Informatics.

[10] Kong, J., Wu, Z.Y., Ismail, M., Serpedin, E. and Qaraqe, K.A., 2019. Q-Learning Based Two-Timescale Power Allocation for Multi-Homing Hybrid RF/VLC Networks. IEEE Wireless Communications Letters.

[11] Zhou, F., Feng, L., Yu, P. and Li, W., 2015, March. Energy-efficiency driven load balancing strategy in LTE-WiFi interworking heterogeneous networks. In 2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW) (pp. 276-281). IEEE.

[12] Lian, J. and Brandt-Pearce, M., 2016, December. Adaptive M-PAM for multiuser MISO indoor VLC systems. In 2016 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6). IEEE.

[13] Li, X., Zhang, R. and Hanzo, L., 2015. Cooperative load balancing in hybrid visible light communications and WiFi. IEEE Transactions on Communications, 63(4), pp.1319-1329.

[14] Alenezi, A.M. and Hamdi, K., 2019, November. Global Q-Learning Approach for Power Allocation in Femtocell Networks. In International Conference on Intelligent Data Engineering and Automated Learning (pp. 220-228). Springer, Cham.

[15] Heegard, C., 2001, September. Range versus rate in IEEE 802.11 g wireless local area networks. In September meeting IEEE (Vol. 802).