IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 2, 2021, accepted February 18, 2021, date of publication February 23, 2021, date of current version March 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3061406

Q-Learning-Based Power Allocation for Secure Wireless Communication in UAV-Aided Relay Network

SIDQY I. ALNAGAR, ANAS M. SALHAB, (Senior Member, IEEE), AND SALAM A. ZUMMO, (Senior Member, IEEE)

Department of Electrical Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Corresponding author: Anas M. Salhab (Salhab@kfupm.edu.sa)

This work was supported by the Deanship of Scientific Research in King Fahd University of Petroleum and Minerals under Grant DF181013.

ABSTRACT Unmanned aerial vehicle (UAV)-aided wireless relay networks are at risk of eavesdropping activities due to their open nature. In this paper, we study the security of a UAV-aided selective relaying wireless network in which N UAVs are employed as decode-and-forward (DF) relays linking a ground base station (BS) with L legitimate users on the ground in the presence of a passive eavesdropper (*Eave*). Direct links between the ground BS and both the ground users and the eavesdropper are assumed to be blocked. The ground-to-air and air-to-ground channels are assumed to follow Rician fading model with opportunistic scheduling scheme for UAVs and users selection. In order to secure data transmissions against such an interception action, the UAV of the worst UAV-selected user link transmits a jamming artificial noise (AN) signal to degrade *Eave* ability in decoding the confidential information successfully. The transmission outage probability, intercept probability, and hybrid outage probability are derived and analyzed. Due to the heavy computation burden raised by increasing the number of UAVs and users as well as the difficulty in estimating the instantaneous channel state information (CSI), existing traditional optimization methods are not highly efficient in solving the considered power allocation problem. Therefore, we propose a dynamic power control scheme based on Q-learning algorithm combined with statistical CSI where the hybrid outage probability is minimized. Simulation results show that the proposed algorithm efficiently reduces the hybrid outage probability with a noticeable reduction in the computational time.

INDEX TERMS Unmanned aerial vehicle, Rician fading, physical layer security, outage probability, intercept probability, reinforcement learning, Q-learning, power allocation.

I. INTRODUCTION

Utilization of unmanned aerial vehicle (UAV) in modern wireless networks is expected to increase rapidly in the next few years. They can serve as aerial base stations (BSs) or relays providing successful solutions to reinforce the network reliability and capacity. UAV-aided networks surpass terrestrial networks in fast deployment and more flexible construction. They can succeed in rapid foundation of supplement wireless network to support terrestrial networks in congested hotspots circumstances. Also, they help in supplying communications in post disaster conditions where the terrestrial network breaks down and communication is demanded by rescue activities [1]. Unlike ground wireless communications, line-of-sight (LOS) is dominant in the UAV-aided networks.

UAV-aided networks can be applied in temporarily events like sports, outdoor activities, and scientific missions [1]. In particular, UAV employment for providing wireless services has enticed wide research and industry efforts in terms of deployment, navigation, and control issues [2]–[4]. Due to the fact that UAVs work in an exposed environments, UAV-aided wireless networks are threatened by eavesdropping attempts from unauthorized parties. Nevertheless, resource allocation such as transmit power is also essential to further enhance the physical layer security (PLS) and outage performance for UAV-aided wireless networks.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiankang Zhang.

TABLE 1. List of acronyms.

Acronym	Description
2D	2 Dimensional
3D	3 Dimensional
A2G	Air-to-ground
AF	Amplify-and-forward
AN	Artificial noise
AWGN	Additive white Gaussian noise
BS	Base station
CDF	Cumulative distribution function
CSI	Channel state information
D2D	Device-to-device
DF	Decode-and-forward
e2e	End-to-end
EH	Energy harvesting
FDD	Frequency division duplex
LOS	Line-of-sight
NLOS	Non-line-of-sight
PLOS	Probability of line-of-sight
PLS	Physical layer security
SINR	Signal-to-interference-plus-noise ratio
SNR	Signal-to-noise ratio
UAV	Unmanned aerial vehicle

To improve UAV-aided network performance in terms of transmission outage and security outage probabilities, a substantial effort has been recently spent on optimization theory to develop more efficient algorithms to gain optimal UAV trajectory or power allocation. However, many works found in literature assume a UAV working in static environment. Practically, in UAV-aided wireless networks with a massive number of users, the environment is often dynamically changeable. Therefore, it is desirable to enable the network controller to have autonomous decision based on local observations, e.g., nodes power, number of network nodes, and their locations.

UAVs employment as relay has recently attracted extensive research efforts. In [5], a UAV relay network was investigated where derivation of the system outage probability was provided. Additionally, a variable rate protocol for a UAV relay hovering at a constant height around a circular path was proposed where the data-rate was optimized to improve the system outage probability. Energy harvesting (EH) applications in UAV-based relay networks were addressed in [6]. The UAV was used to link two ground nodes where air-to-ground (A2G) and ground-to-air (G2A) channels were assumed to follow Rician fading model. The system outage probability was derived and analyzed. Also, the impact of the UAV height on the outage probability were addressed. Another study was achieved in [7] where both the spectrum and energy efficiency optimizations were conducted. The trade-off between these two important metrics was studied for a UAV-based relay network.

The authors in [8] proposed an optimization scheme for the outage probability of a UAV-based relay network by designing the trajectory and controlling the transmit power of the UAV. In [9], the system total throughput was addressed for a hovering UAV where the authors proposed a framework to jointly optimize the power allocation and the UAV trajectory to maximize the system throughput. The scenario of multiple users was not considered in this work. In [10], an optimization of the UAV altitude to improve a UAV relaying system performance in terms of outage probability, bit error rate, and power loss was conducted. The authors in [11] proposed a 3D location optimization scheme to improve the UAV-based relay network outage performance.

A UAV relay was used to link a BS to multiple users where a mixed Rayleigh/Rician channel was assumed. Most of the available papers, which studied the PLS issues in UAV-based relay networks are mainly focusing on optimizing the UAV trajectory or power allocation to enhance the security performance or using jamming techniques to decrease the possibility of eavesdropping. In [12], the secrecy rate of a UAV transmitting information to a ground user was studied in existence of an eavesdropper. The secrecy rate was maximized by jointly optimizing the UAV 2D trajectory and power where the A2G channels were assumed to be dominated by the LOS model. The authors of [13] have addressed the performance of a UAV-based relay connecting two points on the ground in the presence of an eavesdropper in a known location on the ground. LOS channel model was assumed for the A2G links, both the UAV 2D trajectory and power allocation were optimized to maximize the secrecy energy efficiency. In [14], the secrecy rate was maximized by optimizing the power allocation in a UAV-based relay network where free space path loss model was used to model the link between a frequency division duplex (FDD) buffer UAV relay and two ground BSs (source and destination).

In [15], a UAV was used to communicate with multiple destinations on the ground and another UAV was used as a jammer. Joint power and trajectory optimization was achieved with an objective to maximize the secrecy rate. A2G channels were modeled using the LOS model and frequency division multiple access was utilized for multiple destinations scheduling. In [16], a UAV was utilized as a friendly jammer to protect information against potential wiretapping where the information is transmitted between two points on the ground. The source and destination nodes communicate via a Rayleigh channel, while the A2G channel is modeled using the LOS/NLOS probability model. The UAV position and jamming power were optimized to minimize both the outage probability at the legitimate user and the interception probability at the eavesdropper. Cooperative jamming was studied in [17] and [18] where in [17], a transmission scheme was proposed for a UAV-based relay network. The system model contained one destination with the A2G channel modeled by Rician distribution. A transmission scheme that combines simultaneous wireless information and power transfer (SWIPT) energy harvesting approach and

cooperative jamming was also proposed. In addition, exact and asymptotic expressions for the connection and interception probabilities were derived. In [18], the performance of a multiple UAVs relay network was studied where the best UAV is used to relay confidential information, while the remaining UAVs emit jamming signals to protect the information from eavesdropping. Rician fading distribution was used to model the A2G links. Another study was conducted in [19] with a UAV relay connecting a source to a destination in a mmWave scenario with SWIPT technique. A Homogeneous Poisson Process distribution of multiple eavesdroppers was assumed, and the A2G channels were assumed to be Nagakami-*m* distributed. Closed-form expressions for the average secrecy rate and the average energy coverage were derived for both the amplify-and-forward (AF) and the decode-and-forward (DF) relaying schemes. Selective UAV-based relaying network was studied in [20] where multiple UAV-aided relays cooperatively connect a ground BS to a ground destination in existence of one UAV eavesdropper. One UAV is selected opportunistically for forwarding the signal from the BS to the destination, while the remaining UAVs emit jamming signal to degrade the eavesdropper ability to successfully decode the confidential information. In this paper, both the A2G and the G2A channels are modeled using Rician fading model. The authors derived and analyzed the outage probabilities at both the eavesdropper and destination sides. The PLS issue was also addressed in [21] and [22] where a single UAV and a single user model was assumed in [21]. The power allocation was optimized to minimize the system intercept probability. A jamming UAV was used in [22] where the author proposed a system consists of a single UAV-enabled relay connecting a source with a single user on the ground and a single UAV jammer in the presence of a passive eavesdropper. Joint power and trajectory optimization of the two UAVs was achieved to minimize the system intercept probability.

Due to the fact that modern wireless networks are characterized by fast dynamic change, some recent researchers introduced the Reinforcement Learning (RL) to solve the wireless networks optimization problem such as power and spectrum allocation problems with tolerable time delay. Q-learning, which is a model-free RL technique, was applied to solve the power and spectrum allocation problems in different wireless network structures [23]–[25]. Dynamic resource allocation has been provided in [26]–[28] where RL methods were adopted to solve the optimization problems related to the transmit power levels and spectrum allocations. In [26], the scenario of multiple UAVs acting as aerial BSs serving and multiple users on the ground was studied where the A2G channels were modeled using the LOS and the probabilistic models. The authors in [27] addressed a joint power and spectrum allocation optimization problem for multiple UAVs working as aerial BSs. The objective function considered both the propulsion power and signal-to-interference-plus-noise ratio (SINR) requirements of the served users. In [28], the trajectory and power of a UAV employed as a cellular BS were optimized to maximize the sum-rate of vehicle-to-cellular (V2C) communications.

As can be noticed, none of the previous papers addressed or studied the PLS issue in multiple UAV-based relay networks with multiple users. The considered system can be found in several applications such as in gathering information from ground sensors and in providing communication for outdoor activities such as scientific and rescue missions. In addition, such scenario can be used in providing communication services for outdoor events such as in sport activities.

To the best of our knowledge, UAV relay network security enhancement using machine learning techniques has not been addressed yet. In addition, considering multiple UAVs relay serving multiple users and assuming Rician channel model in the presence of a passive eavesdropper has not been studied in the available literature. The following points represent a summary of the main contributions of this paper.

- Deriving expressions for the end-to-end (e2e) outage, intercept, and hybrid outage probabilities for the multiple UAV relay network with multiple users over Rician fading channels.
- Analyzing the security performance of the considered model.
- Proposing an optimization scheme for the power allocation to minimize the hybrid outage probability of the considered system.
- Introducing a Q-learning based algorithm to solve the power allocation problem of the considered system.

The rest of this paper is organized as follows. In Section II, the system and channel models are provided. Problem formulation and derivations of the performance measures are presented in Section III. Explanation of the Q-Learning method and the proposed power allocation algorithm based on Q-Learning are presented in Section IV. Simulation results are demonstrated in Section V. Finally, the paper is concluded in Section VI.

II. SYSTEM AND CHANNEL MODELS

The scenario considered in this paper consists of N UAVs operating as selective relays between a ground BS and L users. We denote the set of N UAVs by $\mathcal{M}$, and the set of L users by $\mathcal{L}$, we assume that a passive eavesdropper *Eave* exists in the region. The L users and *Eave* are assumed to exist in an area on the ground, as shown in Figure 1. It is assumed that direct wireless communication between the ground nodes and the ground BS cannot be achieved due to natural or artificial obstacle. This model can be found in an urban or rural areas where the links between the ground BS and the ground nodes (the users and the eavesdropper) are assumed to be blocked by high buildings or natural obstacles like small mountains.

The users and the eavesdropper are assumed to be located in a specific target area where the users represent the legitimate users of the network, while the eavesdropper represents a party prevented from receiving the transmitted

TABLE 2. List of symbols.

Symbol	Description
α	Learning rate
β	Discounting rate
η	Free space path loss exponent
γ	Instantaneous SNR
$\gamma_{\mathrm{g},\mathrm{u}_m}$	Ground BS to the m^{th} UAV link SNR
γ_{out}	SNR threshold
$\gamma_{\mathrm{u}_m,l}$	The m^{th} UAV to the l^{th} user link SNR
$d_{\mathrm{g},\mathrm{u}_m}$	BS to the m^{th} UAV distance
$d_{\mathrm{u}_m,l}$	The m^{th} UAV to the l^{th} user distance
ϵ	Greedy rate
$h_{\mathrm{g},\mathrm{u}_m}$	Bs to the m^{th} UAV channel coefficient
$h_{\mathrm{u}_m,l}$	m^{th} UAV to l^{th} user channel coefficient
K	Rician factor
L	Number of users
$\mathcal{L}$	Set of users
$\mathcal{M}$	Set of UAVs
$n_{\mathrm{g},\mathrm{u}_m}$	Ground BS to m^{th} UAV AWGN signal
$n_{\mathrm{u}_m,l}$	m^{th} UAV to l^{th} user AWGN signal
N_0	AWGN noise power
P_{g}	Power transmitted by the ground BS
P_{J}	Power transmitted by the jamming UAV
P_{u}	Power transmitted by the UAV
P_{int}	Intercept probability
P_{ho}	Hybrid outage probability
P_{out}	Outage probability
Q_1	First order Marcum Q-function
$\Re$	The spectral efficiency
T_i	Number of training episodes
T_s	Number of iterations per episode
$\mathcal{U}$	Decoding set
x_{g}	Sybmol transmitted at the ground BS
$x_{\mathrm{u}_m,l}$	Symbol transmitted at the m^{th} UAV to the l^{th} user

information. In most practical cases, the accurate location and channel state information (CSI) of a passive eavesdropper are unknown to the network [29]. Hence, it is more practical to design the cooperative jamming to reduce the eavesdropper intercept probability in all possible locations within the target area where eavesdroppers are expected to be located in. A real-world example for our proposed model can be found in a work site containing several locations far away from a main control office. The main control office represents the ground BS and the users and the eavesdropper are located in several locations within an area, which we call the target area. The BS location is assumed to be secured enough against eavesdropping, while the target area is assumed to be far away from the main control office and wide such that eavesdropping activities cannot be avoided.

The channels between the ground nodes and the UAVs are assumed to follow the Rician fading model. Each UAV is assumed to hold a DF relay. Although the AF relays are simply implemented, DF shows superior performance [10]. The considered system is applicable in several wireless schemes, including wireless sensor networks, cellular networks, cognitive radio networks, etc.

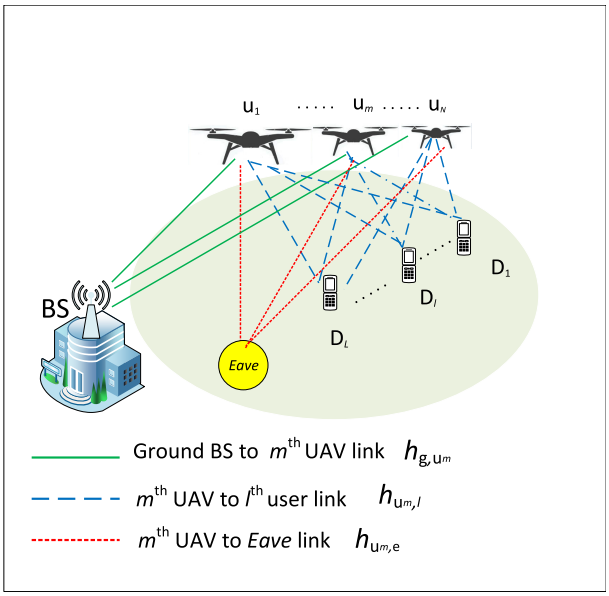


FIGURE 1. UAV relay system model.

During the communication process, the BS sends its signal, where the N UAVs try to decode it. Those UAVs, which achieved the BS signal correctly decoded are put in a decoding set $\mathcal{U}$. Given N UAVs, there exists $2^N - 1$ possible decoding sets given by

$$\Omega = \{\emptyset, \mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_n, \ldots, \mathcal{U}_{2^N-1}\}, \tag{1}$$

where $\emptyset$ represents the case when neither of the N UAVs succeeded in decoding the BS signal, whereas $\mathcal{U}_n$ is a nonempty set from the N UAVs. If the decoding set $\mathcal{U}$ is not empty, a specific UAV-user pair is opportunistically selected for communication, whereas, the worst UAV, in terms of SNR with respect to the selected user, transmits an AN jamming signal to degrade the ability of *Eave* to decode the confidential signals. The AN signal in our paper is assumed to be independent of the information signal and a priori known by the legitimate receiver. This assumption has been used before in several papers in literature [30]–[32]. The AN signal can be easily removed from the confidential information using a self-interference canceling (SIC) receiver [33].

The proposed system model depends on statistical channel modeling instead of instantaneous CSI. Considering instantaneous CSI requires that the receiver continuously estimates the CSI by decoding a symbol transmitted by the transmitter and then feeding back the CSI to the transmitter to adjust accordingly. This process introduces more complexity, higher cost, and higher power consumption.

A. GROUND BS-TO-UAV CHANNEL

The received signal at the m^{th} UAV is given by

$$y_{\mathrm{g},\mathrm{u}_m} = \sqrt{P_{\mathrm{g}}} h_{\mathrm{g},\mathrm{u}_m} x_{\mathrm{g}} + n_{\mathrm{g},\mathrm{u}_m}, \tag{2}$$

where P_{g} is the power transmitted by the ground BS, $h_{\mathrm{g},\mathrm{u}_m}$ is the BS to the m^{th} UAV channel coefficient, which is modeled using Rician fading distribution with $\mathbb{E}[|h_{\mathrm{g},\mathrm{u}_m}|^2] = 1$, where $\mathbb{E}[.]$ is the average operator, x_{g} is the data symbol transmitted by the ground BS with $\mathbb{E}[x_{\mathrm{g}}^2] = 1$, and $n_{\mathrm{g},\mathrm{u}_m} \sim \mathcal{N}(0, N_0)$ is the additive white Gaussian noise (AWGN) term with zero mean and variance N_0.

The instantaneous signal-to-noise ratio (SNR) for the channel between the ground BS and the m^{th} UAV is given by

$$\gamma_{\mathrm{g},\mathrm{u}_m} = \frac{P_{\mathrm{g}}|h_{\mathrm{g},\mathrm{u}_m}|^2}{N_0 d_{\mathrm{g},\mathrm{u}_m}^{\eta}}, \tag{3}$$

where η is the path loss coefficient and $d_{\mathrm{g},\mathrm{u}_m}$ is the distance between the ground BS and the m^{th} UAV.

B. UAV-TO-USERS CHANNEL

The signal sent by the m^{th} UAV and received by the l^{th} user is given by

$$y_{\mathrm{u}_m,l} = \sqrt{P_{\mathrm{u}}} h_{\mathrm{u}_m,l} x_{\mathrm{u}_m,l} + n_{\mathrm{u}_m,l}, \tag{4}$$

where P_{u} is the UAV transmit power, $m \in \{1, 2, \ldots, N\}$, $h_{\mathrm{u}_m,l}$ is the UAV to the l^{th} user channel coefficient, $l \in \{1, 2, \ldots, L\}$, which is modeled using Rician distribution with $\mathbb{E}[|h_{\mathrm{u}_m,l}|^2] = 1$, $x_{\mathrm{u}_m,l}$ is the data symbol transmitted by the m^{th} UAV to the l^{th} user with $\mathbb{E}[x_{\mathrm{u}_m,l}^2] = 1$, and $n_{\mathrm{u}_m,l} \sim \mathcal{N}(0, N_0)$ is the AWGN noise term.

The SNR of the m^{th} UAV to the l^{th} user link is given by

$$\gamma_{\mathrm{u}_m,l} = \frac{P_{\mathrm{u}}|h_{\mathrm{u}_m,l}|^2}{N_0 d_{\mathrm{u}_m,l}^{\eta}}, \tag{5}$$

where $d_{\mathrm{u}_m,l}$ represents the distance from the m^{th} UAV to the l^{th} user. We assume independent non-identically distributed (i.n.d.) wireless paths between the N UAVs and the L users. For user selection, opportunistic scheduling is employed, where the UAV-user link with the best instantaneous SNR γ^* is given the opportunity to communicate, γ^* is given by

$$\gamma^* = \max_{\substack{\mathcal{U}_i \in \mathcal{U}_n \\ c \in \mathcal{L}}} \{\gamma_{\mathrm{u}_i,c}\} \tag{6}$$

III. PERFORMANCE ANALYSIS

A. OUTAGE PROBABILITY

According to previous illustration, we can arrive at the outage probability of the considered model using the total probability theorem [34] as

$$P_{\mathrm{out}} = \Pr(\mathcal{U} = \emptyset) + \sum_{n=1}^{2^N-1} \Pr(\mathcal{U} = \mathcal{U}_n) \Pr \times \Big(\max_{\mathrm{u}_i \in \mathcal{U}_n} \gamma_{\mathrm{u}_i,l} \leq \gamma_{\mathrm{out}} \Big), \tag{7}$$

where $\Pr(.)$ is the probability operation and γ_{out} is the outage SNR threshold, which is given by $\gamma_{\mathrm{out}} = 2^{\Re} - 1$, where $\Re$ is the spectral efficiency. The event $(\mathcal{U} = \emptyset)$ represents the case where neither one of the N UAVs succeed in decoding the signal transmitted by the BS, so we can write this event as

$$\Pr(\mathcal{U} = \emptyset) = \Pr\Big(\max_{\mathrm{u}_m \in \mathcal{M}} \gamma_{\mathrm{g},\mathrm{u}_m} \leq \gamma_{\mathrm{out}} \Big). \tag{8}$$

Substituting (3) into (8) yields

$$\Pr(\mathcal{U} = \emptyset) = \prod_{m=1}^{N} \Pr\big(|h_{\mathrm{g},\mathrm{u}_m}|^2 \leq \gamma_1\big), \tag{9}$$

where $\gamma_1 = \frac{\gamma_{\mathrm{out}} N_0 d_{\mathrm{g},\mathrm{u}_m}^{\eta}}{P_{\mathrm{g}}}$. The probability in (9) represents the cumulative distribution function (CDF) of the chi-square random variable $|h_{\mathrm{g},\mathrm{u}_m}|^2$ evaluated at γ_1, using [35, Eq. (8)] and after doing some mathematical manipulations, (9) can be written as

$$\Pr(\mathcal{U} = \emptyset) = \prod_{m=1}^{N} \left[1 - Q_1(\sqrt{2K}, \sqrt{2(K+1)\gamma_1})\right], \tag{10}$$

where Q_1 is the first order Marcum Q-function and K is the Rician factor, which is the ratio of the LOS component power to the NLOS component power. The probability $\Pr(\mathcal{U} = \mathcal{U}_n)$ can be written as

$$\Pr(\mathcal{U} = \mathcal{U}_n) = \prod_{\mathrm{u}_i \in \mathcal{U}_n} \Pr(\gamma_{\mathrm{g},\mathrm{u}_i} \geqslant \gamma_{\mathrm{out}}) \times \prod_{\mathrm{u}_k \in \widetilde{\mathcal{U}_n}} \Pr(\gamma_{\mathrm{g},\mathrm{u}_k} \leqslant \gamma_{\mathrm{out}}), \tag{11}$$

where $\widetilde{\mathcal{U}_{\mathrm{n}}} = (\mathcal{M} - \mathcal{U}_{\mathrm{n}})$ is the complement of $\mathcal{U}_n$. Using the same steps in deriving (10), we get the outage probability as

$$\begin{aligned} P_{\mathrm{out}} = {} & \prod_{m=1}^{N} \left[1 - Q_1(\sqrt{2K}, \sqrt{2(K+1)\gamma_1})\right] \\ & + \sum_{n=1}^{2^N-1} \prod_{\mathrm{u}_i \in \mathcal{U}_n} Q_1(\sqrt{2K}, \sqrt{2(K+1)\gamma_2}) \\ & \times \prod_{\mathrm{u}_k \in \widetilde{\mathcal{U}_n}} \left[1 - Q_1(\sqrt{2K}, \sqrt{2(K+1)\gamma_3})\right] \\ & \times \prod_{\substack{\mathrm{u}_i \in \mathcal{U}_n \\ l \in \mathcal{L}}} \left[1 - Q_1(\sqrt{2K}, \sqrt{2(K+1)\gamma_4})\right], \end{aligned} \tag{12}$$

where $\gamma_2 = \frac{\gamma_{\mathrm{out}} N_0 d_{\mathrm{g},\mathrm{u}_i}^{\eta}}{P_{\mathrm{g}}}$, $\gamma_3 = \frac{\gamma_{\mathrm{out}} N_0 d_{\mathrm{g},\mathrm{u}_k}^{\eta}}{P_{\mathrm{g}}}$, and $\gamma_4 = \frac{\gamma_{\mathrm{out}} N_0 d_{\mathrm{u}_i,l}^{\eta}}{P_{\mathrm{u}}}$.

B. INTERCEPT PROBABILITY

According to the proposed cooperative jamming scheme, the best UAV-user link will be engaged in the communication process at a time instance, while the UAV with the worst SNR to the selected user is to be used as a jammer emitting a jamming AN signal, which is known to the receiving ends. The power of the jamming signal is denoted by P_{J}. The worst UAV is selected as a jammer to reduce any interference caused by the jammer on the ground user in the case of imperfect CSI [36]. For nonempty subset $\mathcal{U}_n$ and using the law of total probability, the considered system intercept probability is

given by

$$P_{\text{int}} = \sum_{n=1}^{2^N-1} \Pr(\mathcal{U} = \mathcal{U}_n) \sum_{l=1}^{L} \sum_{i=1}^{|\mathcal{U}_n|} \Pr\Big(\max_{\substack{(j,c)\neq(i,l) \\ c\in\mathcal{L}}} \gamma_{\mathrm{u}_j,c} \leq \gamma_{\mathrm{u}_i,l} \Big) \times \Pr\left(\gamma_{\mathrm{u}_i,e} \geqslant \gamma_{\text{out}}\right) \Pr\Big(\min_{\mathrm{u}_t \in \{\mathcal{U}_n - \mathrm{u}_i \& \mathrm{u}_p\}} \gamma_{\mathrm{u}_t,l} > \gamma_{\mathrm{u}_p,l} \Big), \tag{13}$$

where $|\mathcal{U}_n|$ is the cardinality of the set $\mathcal{U}_n$, $\gamma_{\mathrm{u}_i,e}$ is the instantaneous SNR of the selected UAV-to-*Eave* link, i and p are indexes of the transmitting and jamming UAVs respectively, and $t \in \{1, 2, \ldots, N\} - \{i, p\}$. The derivation of the intercept probability P_{int} is provided in Appendix VI. Accordingly, P_{int} is given by

$$\begin{aligned} P_{\text{int}} = & \sum_{n=1}^{2^N-1} \prod_{\mathrm{u}_i \in \mathcal{U}_n} Q_1(\sqrt{2K}, \sqrt{2(K+1)\gamma_2}) \\ & \times \prod_{\mathrm{u}_k \in \widetilde{\mathcal{U}}_n} \left[1 - Q_1(\sqrt{2K}, \sqrt{2(K+1)\gamma_3})\right] \\ & \times \sum_{l=1}^{L} \sum_{i=1}^{|\mathcal{U}_n|} \sum_{v=0}^{|\mathcal{U}_n - \mathrm{u}_i|} \frac{(-1)^v}{(v!)} \sum_{n1=1}^{|\mathcal{U}_n - \mathrm{u}_i|} \sum_{p1=1}^{|\mathcal{U}_n - \mathrm{u}_i|} \cdots \\ & (nx, py) \neq (i, l) \cdots \sum_{nv=1}^{|\mathcal{U}_n - \mathrm{u}_i|} \sum_{pv=1}^{|\mathcal{U}_n - \mathrm{u}_i|} \sum_{s=0}^{\infty} \\ & \frac{2e^{-2K} K^s (K+1)^s (s!)^2 \sigma_{\mathrm{u}_i,l}^{(s+1)} (s+1)!}{\left(\sigma_{\mathrm{u}_i,l} + \sum_{t=1}^{|\mathcal{U}_n - \mathrm{u}_i|} \sigma_{\mathrm{u}_{nt},pt}\right)^{(s+2)}} \\ & \times \sum_{r=0}^{M} \sum_{j=0}^{r} \sum_{s=0}^{\infty} \frac{2 g_r r! \sigma_e^j K^s e^{-K} (s+j+1)}{j! (s!)^2 (K+1)^s (\sigma_e+1)^{s+j+1}} \\ & \sum_{v=0}^{|\mathcal{U}_n - \mathrm{u}_i - \mathrm{u}_p|} \frac{(-1)^v}{(v!)} \sum_{n1=1}^{|\mathcal{U}_n - \mathrm{u}_i - \mathrm{u}_p|} \cdots v \neq (i, p) \cdots \\ & \sum_{nv=1}^{|\mathcal{U}_n - \mathrm{u}_i - \mathrm{u}_p|} \sum_{s=0}^{\infty} \frac{2e^{-2K} K^s (K+1)^{s+1} \sigma_{\mathrm{u}_p,l}^{(s+1)} (s+1)!}{\left(\sigma_{\mathrm{u}_i,l} + \sum_{t=1}^{|\mathcal{U}_n - \mathrm{u}_i - \mathrm{u}_p|} \sigma_{\mathrm{u}_{nt},l}\right)^{(s+2)}}. \end{aligned} \tag{14}$$

C. HYBRID OUTAGE PROBABILITY

Three mutually exclusively events can occur during the data transmission in the considered UAV-based relaying network, namely the transmission outage event, the secrecy outage event, and the secure transmission event. Specifically, the transmission outage event probability is given by (12), and the secrecy outage event probability is given by (14). The hybrid outage probability P_{ho} is utilized as a comprehensive performance measure, which is the sum of the transmission and the secrecy outage probabilities [37]

$$P_{\text{ho}} = P_{\text{out}} + P_{\text{int}}. \tag{15}$$

D. PROBLEM FORMULATION

To enhance the security performance of the considered system, the power allocation optimization problem is formulated as follows

$$\begin{aligned} \underset{P_{\mathrm{g}}, P_{\mathrm{u}}, P_{\mathrm{J}}}{\text{minimize}} \ & P_{ho}(P_{\mathrm{g}}, P_{\mathrm{u}}, P_{\mathrm{J}}) \\ \text{subject to } & P_{\mathrm{u}} \leq P_{\mathrm{u}}^{max} \\ & P_{\mathrm{g}} \leq P_{\mathrm{g}}^{max} \\ & P_{\mathrm{J}} \leq P_{\mathrm{J}}^{max}, \end{aligned} \tag{16}$$

where P_{u}^{max}, P_{g}^{max}, and P_{J}^{max} are the maximum allowed UAV, BS, and the jammer UAV powers, respectively.

IV. POWER ALLOCATION ALGORITHM BASED ON Q-LEARNING

In this section, we discuss a powerful model-free algorithm called Q-learning, and we also illustrate how Q-learning is related to Reinforcement Learning (RL), which is how to obtain Q-learning from RL [39], [40].

A. Q-LEARNING CONCEPTS

A RL model has 4 parameters, which are: a set of the possible states of the environment, a set of the possible actions the agent may take, scalar reward signal, and the policy. These four features are denoted by S, A, R, and π.

- **System states (S):** It describes the circumstance of the UAV-based relaying system environment, and action decision is taken based on the states of the network. The key factors affecting the state of the network environment are the channel and transmit power of the different nodes in the network, and the number of nodes and their locations. All this information is fed back to the network controller, which is considered as the learner in this learning process, so it can adjust the network accordingly to improve the network performance. The system state S is defined as a countable set as

$$S = S(u, e, p) = \{S_0, S_1, \ldots, S_t, \ldots, S_T\}, \tag{17}$$

where u and e represent the users and *Eave* information, respectively. $p = [P_{\mathrm{g}}, P_{\mathrm{u}}, P_{\mathrm{J}}]$.

- **Action space (A):** The controller takes a decision by observing the state of the network, causing the network to change to the next state. The action in the power allocation case means adjusting P_{g}, P_{u}, and P_{J} levels based on the state of the network. Thus, the set of all actions is expressed as

$$A = A(p) = \{A_0, A_1, \ldots, A_t, \ldots, A_T\}. \tag{18}$$

For the action selection, we adopt the ϵ-greedy policy to take advantages of the exploitation and exploration [38]. Specifically, at the current state S_i, the best action maximizing the action-value function is selected with the probability ϵ, where $\epsilon \in [0, 1]$. On the other hand, with the probability $1 - \epsilon$, the action is randomly chosen from A, and this is done for exploration to ensure obtaining global optimal solution. At an iteration i, the action

will be chosen as follows

$$A_i = \begin{cases} \arg\max_{a \in A} Q(S, A) & \text{with probability } 1-\epsilon, \\ \mathcal{U}(A) & \text{with probability } \epsilon, \end{cases} \tag{19}$$

where $Q(S, A)$ is the state-action function defined in (26) and $\mathcal{U}(A)$ is uniformly distributed over the set A.

- **Reward function (R):** In the considered system, the learner tries to maximize the accumulated rewards by taking a set of actions, which directly results in improvement of the system outage and security performance. In the optimization problem (16), the goal is to minimize the system hybrid outage probability. Thus, we define the immediate reward as the amount of change between the current and previous system hybrid outages. The immediate reward is positive when the outage decreases, otherwise it is negative. Thus, the immediate reward can be given as

$$R_t = P_{\text{ho},t} - P_{\text{ho},(t+1)}, \tag{20}$$

where $P_{\text{ho},t}$ is the hybrid outage probability at the time instant t and $P_{\text{ho},(t+1)}$ is the hybrid outage probability at time instant $(t+1)$ after the controller takes the action A_t to change the network from state S_t to state $S_{(t+1)}$.

- **The policy (π):** When the learner is in state S_t, it can take a certain action $A_t = \pi(S_t)$. The objective of a learner is to find the optimal policy that results in maximization of the total expected reward over the operating time, which is described as

$$V^{\pi}(S_t) = \sum_{i=0}^{T_{max}} \beta^i R_{(t+i)}, \tag{21}$$

where $\beta \in [0, 1]$ is the reward discount factor. If β is set to 0, that means only immediate reward R_t is considered into account, on the other hand, if β is close to 1, that means the future reward is more import than the immediate reward. The optimal policy π^* is the policy that maximize the accumulated reward and it is given as follows

$$\pi^* = \arg\max_{A_t} V^{\pi}(S_t), \ \forall (S_t). \tag{22}$$

Substituting (21) into (22), we get

$$\pi^* = \arg\max_{A_t} \left[R(S_t, A_t) + \beta \, V^{\pi}(S_{t+1}, A_{t+1}) \right]. \tag{23}$$

Getting the optimal policy in (23) requires perfect knowledge of the state-action information, which is quiet difficult, so we define

$$Q(S_t, A_t) = R(S_t, A_t) + \beta \, V^{\pi}(S_{t+1}, A_{t+1}). \tag{24}$$

So that

$$\pi^* = \arg\max_{A_t} Q(S_t, A_t). \tag{25}$$

Algorithm 1 Dynamic Power Allocation Algorithm Based on Q Learning

Input: Q-table, $\alpha \in [0, 1]$, $\beta \in [0, 1]$, $\epsilon \in [0, 1]$, L, N, $S(u, p)$, $A(p)$, T_i, T_s
Output: Optimal strategy π^*, Optimal power levels P^*

```
for t1 = 1 to Ti do
    Select an initial state S0 randomly
    for t2 = 1 to Ts do
        Initialize a random number μ ~ U [0, 1]
        if μ > ε then
            Exploit
            Select an action At based on greedy strategy
            Obtain immediate reward Rt and next state St+1
            Update the Q-table according to (26)
            Adjust the network transmit powers according to At
        else
            Explore
            Select an action At randomly
            Obtain immediate reward Rt and next state St+1
            Update the Q-table according to (26)
            Adjust the network transmit powers according to At
        end if
    end for
end for
```

An iterative method is always taken in computing Q-learning as follows

$$Q(S_t, A_t) = (1-\alpha) Q(S_t, A_t) + \alpha \big[R(S_t, A_t) + \beta \ \max Q(S', A') \big], \tag{26}$$

where $0 < \alpha < 1$ is the learning rate.

B. Q-LEARNING-BASED POWER ALLOCATION ALGORITHM

The details of power allocation algorithm based on Q-learning method is given in Algorithm 1. The parameters related to network and Q-learning are initialized in the step input. The number of training episodes T_i is defined, along with T_s, which determines the maximum iterations per training episode. The learner reads the initial state information S_0 and selects an action according to the ϵ-greedy policy to obtain immediate reward and update the corresponding state-action function $Q(S_t, A_t)$. To achieve switching between exploration and exploitation a number μ is taken randomly from a uniform range $[0, 1]$. If μ is greater than ϵ, then exploitation is achieved, where the action which maximizes the state-action function $Q(S, A)$ is to be selected. Otherwise, exploration process is archived, where an action is randomly selected from A. Exploration helps in increasing the chance of reaching global solutions. The optimal power allocation policy can be reached through massive training iterations. The power allocation scheme is obtained from the state-action

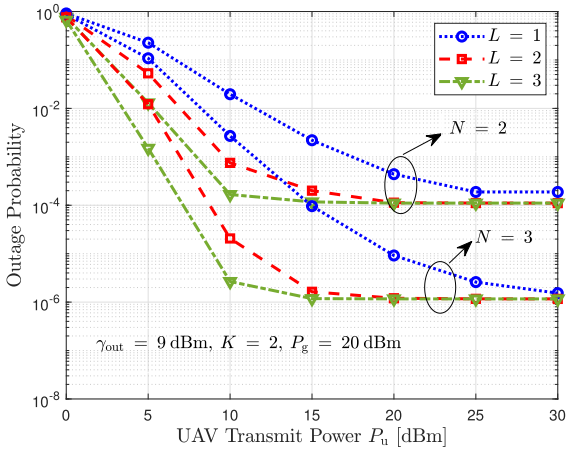


FIGURE 2. Outage probability versus UAV transmit power for different numbers of users and UAVs.

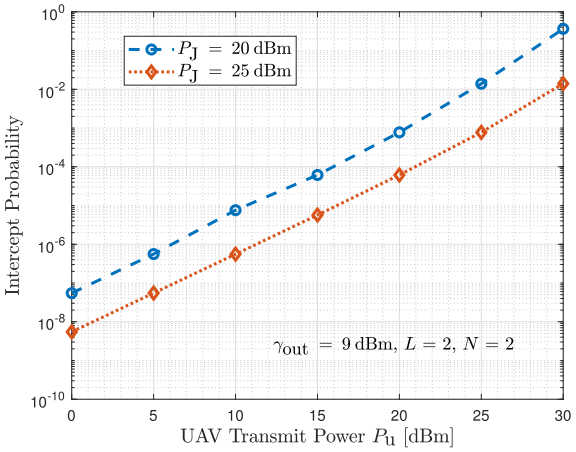


FIGURE 3. Intercept probability versus UAV transmit power for different values of jamming power.

function as follows

$$a_i^* = \arg\max_{a \in A} Q(s_i, a), \quad \forall i. \tag{27}$$

V. SIMULATION AND NUMERICAL RESULTS

In this section, some simulation results are shown and analyzed to validate the derived expressions and the proposed Q-learning based power allocation algorithm. The analytical results of the outage probability and the intercept probability are demonstrated. Moreover, the impacts of various network parameters such as the number of users L, number of UAVs N, UAV transmit power, and the ground BS transmit power on the system performance are investigated.

For the Q-learning based power allocation, we assume the ground users and the eavesdropper are randomly located in a 2D circular area of radius 500 m on the ground. The power levels of the BS, UAV, and the jammer are quantized into three levels (10 dBm, 20 dBm, and 30 dBm). In the same way, the distances from the UAVs to ground nodes are quantized into three levels (100 m, 300 m, and 500 m). This is done to obtain a discrete set of possible states and actions required by the Q-learning algorithm operation. The simulation and experimental parameters are summarized in Table 3.

TABLE 3. Table of simulation and experimental parameters.

Parameter	Value	Comments
Frequency	2000 KHz	Operating frequency
N_0	$1 \times 10^{-8} W / -100$ dB	Thermal noise power
K	2	Rician-K factor
γ_{out}	10 dB	SNR threshold
η	2	Path loss coefficient
α	0.02	Learning rate
β	0.98	Discounting factor
ϵ	0.4	Greedy rate
T_i	10000	Number of training episodes
T_s	400	Iterations per episode

Figure 2 shows the system outage probability versus the UAV transmit power P_u for different numbers of users and UAVs. Two sets of curves are displayed on this figure, set of $N = 2$ and set of $N = 3$. Clearly, when N is high, better performance is obtained, as expected. For both sets, an error floor appears at higher values of P_u as at this range of P_u values, the system performance is dominated by the first hop (fixed P_g) and any increase in either P_u or number of users L adds no gain to the system performance. Increasing P_u leads to improving the system performance when P_g is higher than P_u. In this case also, increasing L adds some gain to the system performance.

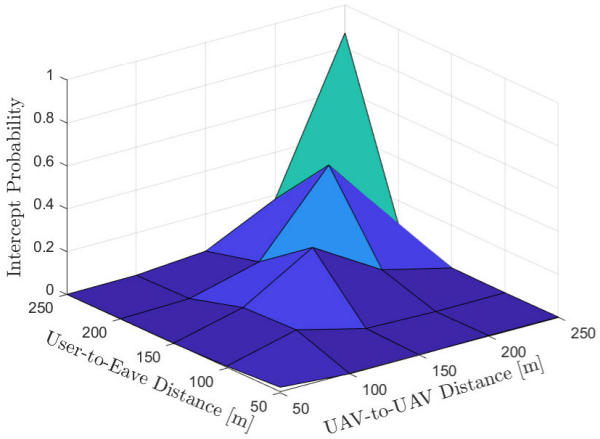


FIGURE 4. Intercept probability versus UAV-to-UAV and user-to-eavesdropper distances, $N = 2$, $L = 1$, $P_g = P_u = P_J = 10$ dBm.

In Figure 3, the jamming power effect on the system intercept probability is demonstrated. As expected, increasing the jamming power P_J results in decreasing the system intercept probability, while increasing P_u has an opposite effect on the intercept probability.

In Figure 4, the intercept outage probability versus the distance between the UAVs and the distance between the user and the eavesdropper is displayed. As can be seen, the distance between the UAVs compared to the distance between the user and the eavesdropper greatly affects the system intercept probability. It is seen that when the distances are comparable, higher intercept probability is resulted.

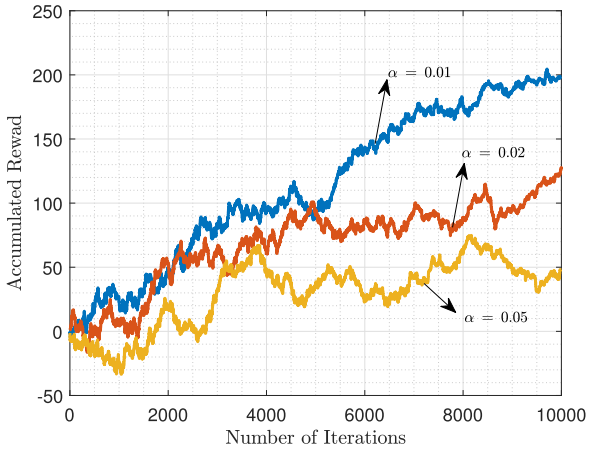


FIGURE 5. Accumulated reward versus the number of iterations for several values of the learning rate α.

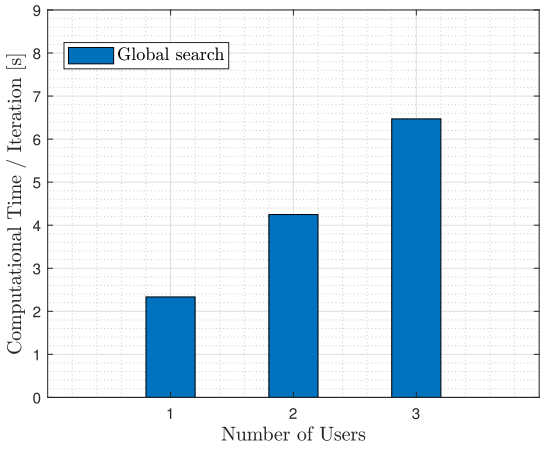


FIGURE 7. Delay time for global search algorithm versus number of users for $N = 3$.

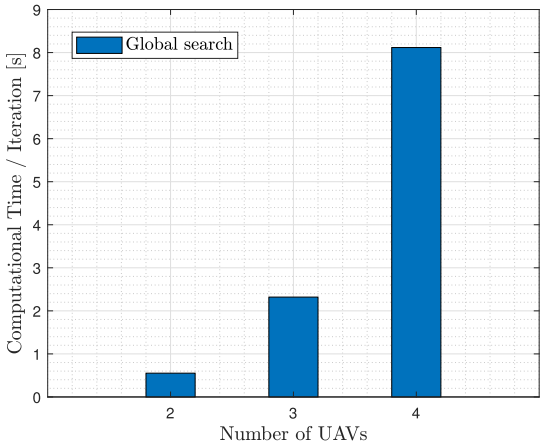


FIGURE 8. Delay time for global search algorithm versus number of UAVs, $L = 1$.

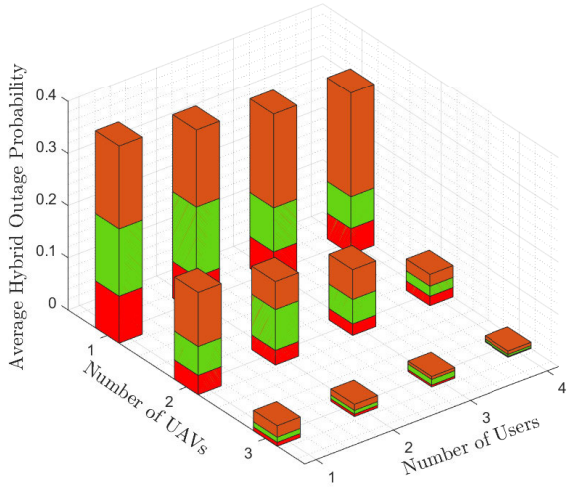


FIGURE 6. Average hybrid outage probability versus number of users and UAVs. The red, green, and brown bars represent the proposed power allocation, the random power allocation, and the maximum power allocation schemes, respectively.

Figure 5 shows the convergence of the accumulated reward under different learning rates α versus number of iterations. The accumulated reward is calculated as the amount of change in the system hybrid outage probability in every iteration. This figure shows that as the number of iterations increases, the accumulated reward gradually converges to a constant value. As α increases, the algorithm requires less number of iterations to converge. The algorithm in this case converges to lower values of the accumulated rewards, which may not lead to learning the optimal strategy. When α is set at smaller value, the algorithm converges to the optimal policy in a slower rate with a higher accumulated reward. So it is required to properly select the value of α to compromise between the speed of convergence and the accumulated reward.

In Figures 6, the hybrid outage probability is displayed versus number of users for different numbers of UAVs. The performance of the proposed algorithm is compared with other two power allocation schemes, namely equal power distribution on all the network nodes (BS, transmitting UAV, and jamming UAV), and random power allocation. The proposed scheme achieves power allocation based on Q-learning where the nodes powers are optimally distributed. The results show the superiority of the proposed algorithm over the other two schemes. By comparing our result with the models proposed in [21] and [22], which can be considered as special cases of the multi-UAV and multi-user model proposed in our paper, we notice that a considerable improvement in the security and outage performance is obtained by the proposed model in this paper.

Global search optimization algorithms such as Genetic algorithm and Particle Swarm provide the optimal global solution in expense of computational time. For dynamic adaptable networks applications, those algorithms are not applicable despite of their high quality solutions compared to the Q-learning based algorithm. The main advantage of the proposed algorithm is its capability to reduce the time required to solve the power allocation problem in a dynamically changing UAV wireless network.

A time performance comparison of the global search methods and the proposed algorithm is presented in Figures 7, 8, 9, and 10. It can be noticed from these figures

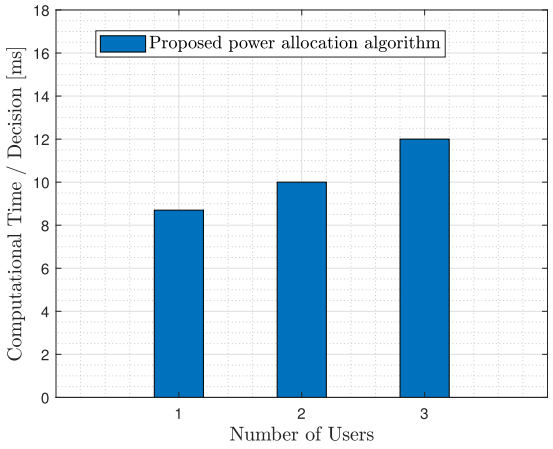


FIGURE 9. Delay time of the proposed algorithm versus number of users, $N = 3$.

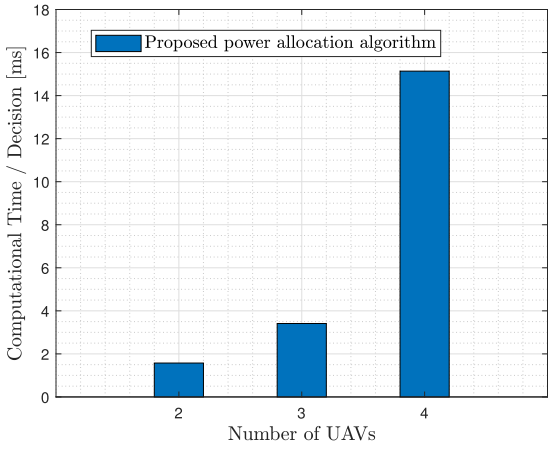


FIGURE 10. Delay time of the proposed algorithm versus number of UAVs, $L = 1$.

that increasing the number of users results in a linear increase in the computation time for both the global search and the Q-learning based algorithm, while increasing the number of UAVs leads to a logarithmic increase in the computation time.

VI. CONCLUSION

In this work, we studied a wireless communication configuration, which utilizes N UAVs as DF opportunistically selected relays. The multiple UAVs were employed to connect the L users to a ground BS where on-ground communication is blocked. The outage, intercept, and hybrid probabilities closed-form expressions were derived. An optimization based on Q-Learning machine learning techniques was proposed to solve the dynamically changing wireless network power allocation problem. The objective of the proposed power allocation algorithm was to minimize the system hybrid outage probability. Our results showed that the proposed algorithm can efficiently reduce the hybrid outage probability compared to equal and random power allocation schemes with a noticeable reduction in the delay time.

APPENDIX

In this appendix, the derivations related to (14) are presented as follows

$$\Pr\Big(\max_{(j,c)\neq(i,l)c\in\mathcal{L}} \gamma_{\mathrm{u}_j,c} \leq \gamma_{\mathrm{u}_i,l}\Big) = \int_0^\infty \prod_{(j,c)\neq(i,l)c\in\mathcal{L}} \Big[1 - Q_1\big(\sqrt{2K}, \sqrt{2(K+1)\sigma_{\mathrm{u}_j,c}x}\big)\Big] \times f_{\gamma_{\mathrm{u}_i,l}}(x)dx. \tag{A.28}$$

According to [42, Eq. (2.10)] the probability density function (PDF) in (A.28) is given by

$$f_{\gamma_{\mathrm{u}_i,l}}(x) = 2(K+1)\sigma_{\mathrm{u}_i,l}e^{-K}xe^{-(K+1)\sigma_{\mathrm{u}_i,l}x} \times I_0\big(2\sqrt{K(K+1)\sigma_{\mathrm{u}_i,l}x}\big), \tag{A.29}$$

where $\sigma_{\mathrm{u}_i,l} = \frac{P_{\mathrm{u}}}{N_0 d^{\eta}_{\mathrm{u}_i,l}}$ and $\sigma_{\mathrm{u}_j,c} = \frac{P_{\mathrm{u}}}{N_0 d^{\eta}_{\mathrm{u}_j,c}}$ and $I_0(.)$ is the zeroth-order modified Bessel function of the first kind, and it is given by

$$I_0(z) = \sum_{s=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^s}{(s!)^2}. \tag{A.30}$$

An approximation of the Marcum Q-function is found in [41, Eq. (7)] as

$$Q_1(W, V) \approx \sum_{r=0}^{M} g_r r! e^{-\frac{V^2}{2}} \sum_{j=0}^{r} \frac{(\frac{V^2}{2})^j}{j!}, \tag{A.31}$$

where M depends on $\max\{1, W, V\}$, which can be truncated as 50 $\max\{1, W, V\}$ [6], and g_r is given by

$$g_r = \frac{\Gamma(1+M)M^{1-2r}W^{2r}2^{-r}}{\Gamma(r+1)\Gamma(M-r+1)\Gamma(1+r)e^{\frac{W^2}{2}}}, \tag{A.32}$$

where $\Gamma(\,.\,)$ is the Gamma function. Upon using (A.31) and (A.32) in (A.28) and using Taylor series expansion of the exponential function with ignoring the higher order terms for high SNR regime, the product term in (A.28) can be simplified as follows

$$\prod_{\substack{(j,c)\neq(i,l)\\ c\in\mathcal{L}}} \Big[1 - Q_1\big(\sqrt{2K}, \sqrt{2(K+1)\sigma_{\mathrm{u}_j,c}x}\big)\Big] = \prod_{\substack{(j,l)\neq(i,l)\\ c\in\mathcal{L}}} \Big[1 - e^{-K}e^{-(K+1)\sigma_{\mathrm{u}_j,c}x}\Big]. \tag{A.33}$$

Given the identity

$$\prod_{v=1}^{V}(1-q_v) = \sum_{v=0}^{V}\frac{(-1)^v}{(v!)}\sum_{n1,\ldots,nv}^{V}\prod_{t=1}^{v}q_{nt}. \tag{A.34}$$

With $\sum_{n1,\ldots,nv}^{V}$ being a short hand of $\sum_{n_1=1}^{} n_1 \neq n_2 \neq .. \neq n_v \sum_{n_v=1}^{}$.

Upon using (A.34), (A.33), (A.32), (A.31), (A.30), and (A.29) into (A.28), then using [43, Eq. (3.361.2)] and [43, Eq. (3.371.1)], we get

$$\begin{aligned}&\Pr\Big(\max_{\substack{(j,c)\neq(i,l)\\ c\in\mathcal{L}}}\gamma_{\mathrm{u}_j,c}\leq\gamma_{\mathrm{u}_i,l}\Big)\\&=\sum_{v=0}^{|\mathcal{U}_n-\mathrm{ui}|}\frac{(-1)^v}{(v!)}\sum_{n1=1}^{|\mathcal{U}_n-\mathrm{u}_i|}\sum_{p1=1}^{|\mathcal{U}_n-\mathrm{u}_i|}\ldots(nx,py)\neq(i,l)\ldots\\&\sum_{nv=1}^{|\mathcal{U}_n-\mathrm{u\_}i|}\sum_{pv=1}^{|\mathcal{U}_n-\mathrm{u}_i|}\sum_{s=0}^{\infty}\frac{2e^{-2K}K^s(K+1)^s(s!)^2\sigma_{\mathrm{u}_i,l}^{(s+1)}(s+1)!}{\left(\sigma_{\mathrm{u}_i,l}+\sum_{t=1}^{|\mathcal{U}_n-\mathrm{u}_i|}\sigma_{\mathrm{u}_{nt},pt}\right)^{(s+2)}}.\end{aligned}\tag{A.35}$$

Using the same procedure, we can get the remaining terms as follows

$$\begin{aligned}&\Pr\Big(\min_{\mathrm{u}_t\in\{\mathcal{U}_n-\mathrm{u}_i\&\mathrm{u}_p\}}\gamma_{\mathrm{u}_t,l}>\gamma_{\mathrm{u}_p,l}\Big)\\&=\sum_{v=0}^{|\mathcal{U}_n-\mathrm{u}_i-\mathrm{u}_p|}\frac{(-1)^v}{(v!)}\sum_{n1=1}^{|\mathcal{U}_n-\mathrm{u}_i-\mathrm{u}_p|}\ldots v\neq(i,p)\ldots\\&\sum_{nv=1}^{|\mathcal{U}_n-\mathrm{u}_i-\mathrm{u}_p|}\sum_{s=0}^{\infty}\frac{2e^{-2K}K^s(K+1)^{s+1}\sigma_{\mathrm{u}_p,l}^{(s+1)}(s+1)!}{\left(\sigma_{\mathrm{u}_i,l}+\sum_{t=1}^{|\mathcal{U}_n-\mathrm{u}_i-\mathrm{u}_p|}\sigma_{\mathrm{u}_{nt},l}\right)^{(s+2)}},\end{aligned}\tag{A.36}$$

and

$$\begin{aligned}&\Pr\left(\gamma_{\mathrm{u}_i,e}\geqslant\gamma_{\mathrm{out}}\right)\\&=1-\Pr\left(\gamma_{\mathrm{u}_i,e}<\gamma_{\mathrm{out}}\right)\\&=1-\Pr\left(\frac{P_{\mathrm{u}}|h_{\mathrm{u}_i,e}|^2d_{\mathrm{u}_i,e}^{-\eta}}{N_0+P_{\mathrm{J}}|h_{\mathrm{u}_p,e}|d_{\mathrm{u}_p,e}^{-\eta}}<\gamma_{\mathrm{out}}\right).\end{aligned}\tag{A.37}$$

At high SNR values, the term N_0 can be ignored, and hence, (A.37) can be rewritten as

$$\begin{aligned}&\Pr\left(\gamma_{\mathrm{u}_i,e}\geqslant\gamma_{\mathrm{out}}\right)\\&=1-\Pr\left(\frac{P_{\mathrm{u}}|h_{\mathrm{u}_i,e}|^2d_{\mathrm{u}_i,e}^{-\eta}}{P_{\mathrm{J}}|h_{\mathrm{u}_p,e}|^2d_{\mathrm{u}_p,e}^{-\eta}}<\gamma_{\mathrm{out}}\right)\\&=1-\Pr\left(|h_{\mathrm{u}_i,e}|^2<\sigma_e|h_{\mathrm{u}_p,e}|^2\right)\\&=1-\int_0^{\infty}[1-Q_1(\sqrt{2K},\sqrt{2(K+1)\sigma_ex})]\\&\quad\times f_{h_{\mathrm{u}_p,e}}(x)dx,\end{aligned}\tag{A.38}$$

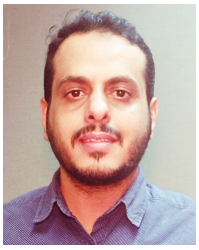
where $\sigma_e = \gamma_{out}(\frac{P_J}{P_u})(\frac{d_{\mathrm{u}_p,e}}{d_{\mathrm{u}_i,e}})^{-\eta}$, d_{u_p},e and $d_{\mathrm{u}_i,e}$ are the distances between the jamming and transmitting UAVs to the eavesdropper, respectively. Using (A.29), (A.30), (A.31), and (A.32) we get

$$\begin{aligned}&\Pr\left(\gamma_{\mathrm{u}_i,e}>\gamma_{\mathrm{out}}\right)\\&=\sum_{r=0}^{M}\sum_{j=0}^{r}\sum_{s=0}^{\infty}\frac{2g_r\,r!\sigma_e^jK^se^{-K}(s+j+1)}{j!(s!)^2(K+1)^s(\sigma_e+1)^{s+j+1}}.\end{aligned}\tag{A.39}$$

REFERENCES

[1] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.
[2] E. W. Frew and T. X. Brown, "Airborne communication networks for small unmanned aircraft systems," *Proc. IEEE*, vol. 96, no. 12, pp. 2008–2027, Dec. 2008.
[3] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart., 2019.
[4] X. Cao, P. Yang, M. Alzenad, X. Xi, D. Wu, and H. Yanikomeroglu, "Airborne communication networks: A survey," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 1907–1926, Sep. 2018.
[5] F. Ono, H. Ochiai, and R. Miura, "A wireless relay network based on unmanned aircraft system with rate optimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7699–7708, Nov. 2016.
[6] L. Yang, J. Chen, M. O. Hasna, and H.-C. Yang, "Outage performance of UAV-assisted relaying systems with RF energy harvesting," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2471–2474, Dec. 2018.
[7] J. Zhang, Y. Zeng, and R. Zhang, "Spectrum and energy efficiency maximization in UAV-enabled mobile relaying," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.
[8] S. Zhang, H. Zhang, Q. He, K. Bian, and L. Song, "Joint trajectory and power optimization for UAV relay networks," *IEEE Commun. Lett.*, vol. 22, no. 1, pp. 161–164, Jan. 2018.
[9] Y. Zeng, R. Zhang, and T. J. Lim, "Throughput maximization for UAV-enabled mobile relaying systems," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 4983–4996, Dec. 2016.
[10] Y. Chen, W. Feng, and G. Zheng, "Optimum placement of UAV as relays," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 248–251, Feb. 2018.
[11] S. I. Alnagar, A. M. Salhab, and S. A. Zummo, "Unmanned aerial vehicle relay system: Performance evaluation and 3D location optimization," *IEEE Access*, vol. 8, pp. 67635–67645, Apr. 2020.
[12] G. Zhang, Q. Wu, M. Cui, and R. Zhang, "Securing UAV communications via joint trajectory and power control," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1376–1389, Feb. 2019.
[13] L. Xiao, Y. Xu, D. Yang, and Y. Zeng, "Secrecy energy efficiency maximization for UAV-enabled mobile relaying," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 1, pp. 180–193, Mar. 2020.
[14] Q. Wang, Z. Chen, W. Mei, and J. Fang, "Improving physical layer security using UAV-enabled mobile relaying," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 310–313, Jun. 2017.
[15] Y. Zhou, P. L. Yeoh, H. Chen, Y. Li, R. Schober, L. Zhuo, and B. Vucetic, "Improving physical layer security via a UAV friendly jammer for unknown eavesdropper location," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11280–11284, Nov. 2018.
[16] X. Zhou, Q. Wu, S. Yan, F. Shu, and J. Li, "UAV-enabled secure communications: Joint trajectory and transmit power optimization," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4069–4073, Apr. 2019.
[17] M. Tatar Mamaghani and Y. Hong, "On the performance of low-altitude UAV-enabled secure AF relaying with cooperative jamming and SWIPT," *IEEE Access*, vol. 7, pp. 153060–153073, Oct. 2019.
[18] R. Ma, W. Yang, Y. Zhang, J. Liu, and H. Shi, "Secure mmWave communication using UAV-enabled relay and cooperative jammer," *IEEE Access*, vol. 7, pp. 119729–119741, Aug. 2019.
[19] X. Sun, W. Yang, Y. Cai, R. Ma, and L. Tao, "Physical layer security in millimeter wave SWIPT UAV-based relay networks," *IEEE Access*, vol. 7, pp. 35851–35862, Mar. 2019.
[20] T. Shen and H. Ochiai, "A UAV-aided selective relaying with cooperative Jammers for secure wireless networks over rician fading channels," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Honolulu, HI, USA, Sep. 2019, pp. 1–5.
[21] G. Sun, N. Li, X. Tao, and H. Wu, "Power allocation in UAV-enabled relaying systems for secure communications," *IEEE Access*, vol. 7, pp. 119009–119017, Sep. 2019.
[22] Y. Li, R. Zhang, J. Zhang, S. Gao, and L. Yang, "Cooperative jamming for secure UAV communications with partial eavesdropper information," *IEEE Access*, vol. 7, pp. 94593–94603, Jul. 2019.
[23] S. Maghsudi and S. Stanczak, "Hybrid centralized–distributed resource allocation for device-to-device communication underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2481–2495, Apr. 2016.

[24] G. Zhao, Y. Li, C. Xu, Z. Han, Y. Xing, and S. Yu, "Joint power control and channel allocation for interference mitigation based on reinforcement learning," *IEEE Access*, vol. 7, pp. 177254–177265, Aug. 2019.
[25] J.-M. Kang, "Reinforcement learning based adaptive resource allocation for wireless powered communication systems," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1752–1756, Aug. 2020.
[26] J. Cui, Y. Liu, and A. Nallanathan, "Multi-agent reinforcement learning-based resource allocation for UAV networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 729–743, Feb. 2020.
[27] Y. Sun, L. Li, Q. Cheng, D. Wang, W. Liang, X. Li, and Z. Han, "Joint trajectory and power optimization in multi-type UAVs network with mean field Q-Learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Dublin, Ireland, Jun. 2020, pp. 1–6.
[28] L. Deng, G. Wu, J. Fu, Y. Zhang, and Y. Yang, "Joint resource allocation and trajectory control for UAV-enabled vehicular communications," *IEEE Access*, vol. 7, pp. 132806–132815, Sep. 2019.
[29] Y. Zhou, P. L. Yeoh, H. Chen, Y. Li, R. Schober, L. Zhuo, and B. Vucetic, "Improving physical layer security via a UAV friendly jammer for unknown eavesdropper location," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11280–11284, Nov. 2018.
[30] H. Long, W. Xiang, J. Wang, Y. Zhang, and W. Wang, "Cooperative jamming and power allocation in three-phase two-way relaying wire-tap systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 4216–4220.
[31] M. L. Jorgensen, B. R. Yanakiev, G. E. Kirkelund, P. Popovski, H. Yomo, and T. Larsen, "Shout to secure: Physical-layer wireless security with known interference," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Washington, DC, USA, Nov. 2007, pp. 33–38.
[32] I. Krikidis, J. S. Thompson, P. M. Grant, and S. McLaughlin, "Power allocation for cooperative-based jamming in wireless networks with secrecy constraints," in *Proc. IEEE Globecom Workshops*, Miami, Fl, USA, Dec. 2010, pp. 1177–1181.
[33] H. Long, W. Xiang, J. Wang, Y. Zhang, and W. Wang, "Cooperative jamming and power allocation with untrusty two-way relay nodes," *IET Commun.*, vol. 8, no. 13, pp. 2290–2297, Sep. 2014.
[34] Y. Zou, J. Zhu, B. Zheng, and Y.-D. Yao, "An adaptive cooperation diversity scheme with best-relay selection in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5438–5445, Oct. 2010.
[35] M. M. Azari, F. Rosas, K.-C. Chen, and S. Pollin, "Ultra reliable UAV communication using altitude and cooperation diversity," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 330–344, Jan. 2018.
[36] A. H. Abd El-Malek, A. M. Salhab, S. A. Zummo, and M.-S. Alouini, "Security-reliability trade-off analysis for multiuser SIMO mixed RF/FSO relay networks with opportunistic user scheduling," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 5904–5918, Sep. 2016.
[37] N. Yang, S. Yan, J. Yuan, R. Malaney, R. Subramanian, and I. Land, "Artificial noise: Transmission optimization in multi-input single-output wiretap channels," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1771–1783, May 2015.
[38] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.
[39] P. Stone and M. Veloso, "Multiagent systems: A survey from a machine learning perspective," *Auto. Robots*, vol. 8, no. 3, pp. 345–383, Jun. 2000.
[40] J. Nie and S. Haykin, "A dynamic channel assignment policy through Q-learning," *IEEE Trans. Neural Netw.*, vol. 10, no. 6, pp. 1443–1455, Nov. 1999.
[41] P. C. Sofotasios and S. Freear, "Novel expressions for the Marcum and one dimensional Q-functions," in *Proc. 7th Int. Symp., Wireless Commun. Syst. (ISWCS)*, York, U.K., Sep. 2010, pp. 736–740.
[42] M. K. Simon and M.-S. Alouini, *Digital Communication Over Fading Channels*, 2nd ed. Hoboken, NJ, USA: Wiley, 2005.
[43] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. New York, NY, USA: Academic, 2007.



SIDQY I. ALNAGAR was born in Aden, Yemen. He received the B.Sc. degree in electronic and communication engineering from the University of Aden, Aden, in 2007. He is currently pursuing the M.Sc. degree in electrical engineering with the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia. His main research interests include unmanned aerial vehicle communication and networking, physical layer security, and wireless communications in general.

ANAS M. SALHAB (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Palestine Polytechnic University, Hebron, Palestine, in 2004, the M.Sc. degree in electrical engineering from the Jordan University of Science and Technology, Irbid, Jordan, in 2007, and the Ph.D. degree from the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 2013. From 2013 to 2014, he was a Postdoctoral Fellow with the Department of Electrical Engineering, KFUPM. He is currently an Associate Professor and the Assistant Director of the Science and Technology Unit with the Deanship of Scientific Research, KFUPM. His research interests include special topics in modeling and performance analysis of wireless communication systems, including cooperative relay networks, cognitive radio relay networks, free space optical networks, visible light communications, and co-channel interference. Recently, he has been nominated for the Excellence Award for Scientific Production (Engineering and Technology Field) by ISI and Scopus among the first ten faculty and researchers in KFUPM, from 2013 to 2018, offered by the Saudi Digital Library, Ministry of Education, Saudi Arabia. In 2014, he was selected as an Exemplary Reviewer by the IEEE Wireless Communications Letters for his reviewing service.

SALAM A. ZUMMO (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 1998 and 1999, respectively, both in electrical engineering, and the Ph.D. degree from the University of Michigan, Ann Arbor, USA, in 2003. He is currently a Professor with the Department of Electrical Engineering, KFUPM. He has authored over 100 papers in reputable journals and conference proceedings and holds six issued U.S. patents. His research interest includes the area of wireless communications, including cooperative diversity, cognitive radio, multiuser diversity, scheduling, MIMO systems, error control coding, multihop networks, and interference modeling and analysis in wireless systems. He was a recipient of the Saudi Ambassador Award for early Ph.D. completion, in 2003, the British Council/BAE Research Fellowship awards, in 2004 and 2006, and the KFUPM Excellence in Research Award, from 2011 to 2012.