Learning-Assisted Secure Relay Selection with Outdated CSI for Finite-State Markov Channel

Jianzhong Lu, Dongxuan He, and Zhaocheng Wang, *Senior Member, IEEE*
Tsinghua University, China

2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring) | 978-1-7281-8964-2/20/$31.00 ©2021 IEEE | DOI: 10.1109/VTC2021-Spring51267.2021.9448708

***Abstract*—In this paper, we investigate secure relay selection for finite-state Markov channel and propose a Q-learning assisted relay selection scheme. Specifically, we firstly analyze the achievable effective secrecy throughput of random selection scheme and optimal selection scheme, respectively, showing that the secrecy performance is highly determined by relay selection methodology. Then, we leverage the Q-learning to learn how to select relay for finite-state Markov channel, which is capable of selecting proper relay with outdated channel state information. Numerical results demonstrate that our proposed Q-learning assisted relay selection scheme can achieve a significant improvement of effective secrecy throughput even with outdated channel information.**

I. Introduction

Due to the broadcast nature of wireless channel, confidential information is vulnerable to be intercepted by illegal receivers, which has to be solved for the next generation of wireless communication. Besides, the decentralized modern wireless networks have introduced challenges to traditional key-based cryptographic techniques, where key generation, distribution, and management are costly and hard to realize. To address this issue, physical layer security has been regarded as an alternative for cryptographic technique to prevent the illegal interception in decentralized networks, which achieves secure transmission without secrecy key [1].

Relay network, as a kind of cooperative decentralized modern wireless network, has been regarded as an efficient wireless network due to its ability of power reduction, coverage extension, and throughput enhancement [2]. In addition, physical layer security in cooperative relay networks has also received much attentions. For example, a secure resource allocation and scheduling for decode-and-forward (DF) relay networks was proposed in [3], which guarantees a non-zero secrecy rate under the requirements of secrecy outage probability and channel outage probability. The authors in [4] provided a relay selection method which realizes a trade-off between the transmission security and reliability for cognitive radio systems. The authors in [5] investigated how to select the secure relay in the presence of untrusted nodes. However, these works have not considered the relay selection in dynamic wireless scenarios, which requires the real-time policy.

Specifically, for dynamic wireless scenarios, the channel state information (CSI) for relay selection is typically outdated due to the delay of channel feedback, thus an efficient relay selection policy that considers the inaccuracy of CSI is required for such scenarios. Recently, researchers have utilized the delay model of channel to characterize the inaccuracy of CSI, and proposed a series of robust relay selection policies based on this model [6]–[8]. For example, two variations of amplify-and-forward (AF) relay selection schemes, namely best relay selection and partial relay selection, were analyzed in [6], where the explicit expressions of outage probability were derived based on the delay model. A cooperative relay selection strategy for DF relay networks was proposed in [7], which realizes robust selection with imperfect CSI. The authors in [8] investigated how to jointly select relay and allocate power in cooperative networks with outdated CSI. Based on the delay model, robust relay selection policies can be obtained, however, such robust relay selection policies cannot follow the change of channel, which degrades the secrecy performance.

Recently, machine learning has been used to solve the complicated physical layer security problems. For example, support vector machine and naive-Bayes have been utilized to facilitate secure transmit antenna selection with low computational complexity in [9], deep neural network has been utilized to optimize secure transmission parameters for wireless powered system in [10], reinforcement learning has been used to realize physical layer security oriented frequency allocation when the backhaul capacity is limited [11]. With the aid of machine learning methods, intractable physical layer security problems might be handled.

In this paper, we study secure relay selection policy for DF relays, where our technical contributions are summarized as follows. Firstly, we analyze the secrecy performance of relay selection for finite-state Markov channel, and evaluate the effective secrecy throughput (EST) for both random selection scheme (RSS) and optimal selection scheme (OSS). Secondly, we exploit the potential of Q-learning in relay selection with outdated CSI, which could obtain a better secrecy performance compared to its random selection counterpart.

II. System Model

As shown in Fig.1, we consider an DF relay wireless network which consists of one source (Alice), one destination (Bob), a passive eavesdropper (Eve), and K relays denoted by $R_1, R_2, \cdots, R_K$. In particular, all nodes are equipped with single antenna. We assume that there is no direct link between Alice and Bob and all the channels are subject to independent and identically distributed (i.i.d) Rayleigh fading. Each time slot can be divided into two phases, where Alice broadcasts information to all the relays with power P_a during the first phase, and the selected relay R_k ($1 \leq k \leq K$) will forward its received information with power P_r during the second phase.

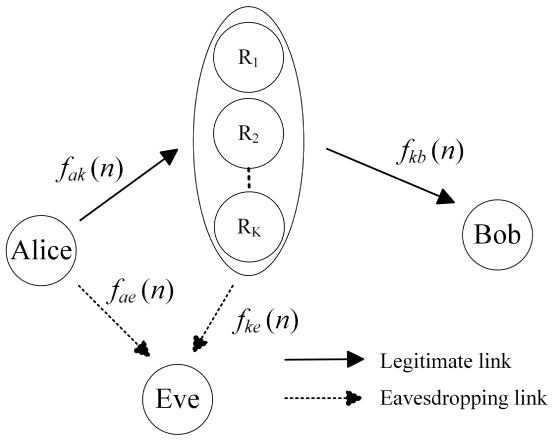

Fig. 1. DF relay system model at n-th time slot.

Specifically, all the legitimate channels are assumed to remain unchanged with one time slot and vary the next time slot, where the channel gains of broadcast channel from Alice to R_k and the relay channel from R_k to Bob at the n-th time slot are expressed as $f_{ak}(n) = |h_{ak}(n)|^2$ and $f_{kb}(n) = |h_{kb}(n)|^2$, respectively, and $h_{ak}(n)$ and $h_{kb}(n)$ denote the channel coefficient at the n-th time slot. Furthermore, Eve can intercept the information form Alice and relays when it locates within the coverage of Alice and all the relays, and the channel gains of channels from Alice and R_k to Eve are denoted as $f_{ae}(n) = |h_{ae}(n)|^2$ and $f_{ke}(n) = |h_{ke}(n)|^2$, which are exponentially distributed random variables with mean value denoted as λ_{ae} and λ_{ke}, and $h_{ae}(n)$ and $h_{ke}(n)$ denote the channel coefficient at n-th time slot, respectively. We also assume that the outdated CSI $f_{ak}(n-1)$ and $f_{kb}(n-1)$ are known at n-th time slot while the statistical information of $f_{ae}(n)$ and $f_{ke}(n)$, i.e. λ_{ae} and λ_{ke} are known to legitimate nodes.

To characterize the dynamic behavior of broadcast channel and relay channel, the finite-state Markov channel model is adopted [12], where $f_{ak}(n)$ and $f_{kb}(n)$ are subject to the Markov process with transition probability $p_{ak}(f_{ak}(n)|f_{ak}(n-1))$ and $p_{kb}(f_{kb}(n)|f_{kb}(n-1))$, respectively. Here, $f_{ak}(n)$ and $f_{kb}(n)$ are elements of $\mathbf{f}_{ak}$ and $\mathbf{f}_{kb}$, respectively, where $\mathbf{f}_{ak} = \left[f_{ak}^1, f_{ak}^2, \cdots, f_{ak}^{M_k}\right]$ and $\mathbf{f}_{kb} = \left[f_{kb}^1, f_{kb}^2, \cdots, f_{kb}^{N_k}\right]$ for $k = 1, \cdots, K$ denote the state spaces of Markov chains for $f_{ak}(n)$ and $f_{kb}(n)$. In addition, the steady-state probability of broadcast channel in state f_{ak}^i and relay channel in state f_{kb}^j are π_{ak}^i for $i = 1, \cdots, M_k$ and π_{kb}^j for $j = 1, \cdots, N_k$, respectively.

Additionally, the signal-to-noise ratio (SNR) of the links Alice-R_k, R_k-Bob, Alice-Eve and R_k-Eve can be expressed as $\gamma_{ak}(n) = P_a f_{ak}(n)/\sigma_k^2$, $\gamma_{kb}(n) = P_r f_{kb}(n)/\sigma_b^2$, $\gamma_{ae}(n) = P_a f_{ae}(n)/\sigma_e^2$, and $\gamma_{ke}(n) = P_r f_{ke}(n)/\sigma_e^2$, respectively, where σ_k^2, σ_b^2, and σ_e^2 are the power of noise at R_k, Bob, and Eve. Without loss of generality, we assume $\sigma_k^2 = \sigma_b^2 = \sigma_e^2 = \sigma^2$.

Once R_k is selected, the SNR of legitimate link for the n-th time slot is given by [13]

$$\gamma_l(n) = \min(\gamma_{ak}(n), \gamma_{kb}(n)). \tag{1}$$

As for Eve, maximal ratio combining (MRC) is adopted to improve its interception ability, where the overall SNR of the eavesdropping link is given by

$$\gamma_e(n) = \gamma_{ae}(n) + \gamma_{ke}(n) = \frac{P_a f_{ae}(n) + P_r f_{ke}(n)}{\sigma^2}. \tag{2}$$

Based on (1) and (2), the total capacity of both legitimate link and eavesdropping link at the n-th time slot can be expressed as

$$C_l(n) = \frac{1}{2}\log_2(1 + \gamma_l(n)), \tag{3}$$

and

$$C_e(n) = \frac{1}{2}\log_2(1 + \gamma_e(n)). \tag{4}$$

In this paper, the well-known Wyner's wiretap code with parameter pair (R_b, R_e) is adopted to guarantee the reliable transmission, where R_b denotes the transmission rate of the wiretap code and R_e denotes the redundancy rate of the wiretap code. Specifically, Bob cannot decode the confidential information successfully when $R_b > C_l(n)$, so the transmission outage probability can be defined as the probability that $C_l(n)$ is below R_b, which can be formulated as

$$P_t^k = \Pr(R_b > C_l(n)). \tag{5}$$

In addition, the information might be leaked to Eve and secrecy outage will happen when $R_e < C_e(n)$, so the secrecy outage probability can be expressed as

$$\begin{aligned} P_s^k &= \Pr(R_e < C_e(n)) \\ &= \frac{\Theta^k}{\lambda_1}\left(e^{-\lambda_1(2^{2R_e}-1)}\right) - \frac{\Theta^k}{\lambda_2}\left(e^{-\lambda_2(2^{2R_e}-1)}\right), \end{aligned} \tag{6}$$

where $\Theta^k = \sigma^2/(P_a\lambda_{ae} - P_r\lambda_{ke})$, $\lambda_1 = \sigma^2/P_a\lambda_{ae}$ and $\lambda_2 = \sigma^2/P_r\lambda_{ke}$.

Proof. See Appendix A. □

To evaluate the transmission reliability and security simultaneously, EST is selected as the performance metric, given by [14]

$$T^k = (R_b - R_e)(1 - P_t^k)(1 - P_s^k), \tag{7}$$

where $R_b - R_e$ represents the target secrecy rate.

We can see that the EST in (7) is determined by the selected relay, where the optimal relay that maximizes EST can be formulated as

$$k^* = \underset{k=1,2,\cdots,K}{\arg\max}\ T^k. \tag{8}$$

III. Performance Analysis

In this section, we investigate the average EST of each relay in the presence of state transition probability for broadcast channel and relay channel and then analyze the average EST of RSS and OSS.

A. Average EST of each relay

Hereby, we analyze the average EST of each relay with certain state transition probability. Firstly, we get the steady-state probability based on the relationship between steady-state probability and its state transition probability, given by

$$(P_{ak}^{M_k})^T Q_{ak}^{M_k} = P_{ak}^{M_k}, \tag{9}$$

which can be expressed as

$$\begin{bmatrix} \pi_{ak}^1 \\ \pi_{ak}^2 \\ \vdots \\ \pi_{ak}^{M_k} \end{bmatrix}^T \begin{bmatrix} p_{ak,11} & \cdots & p_{ak,1M_k} \\ p_{ak,21} & \cdots & p_{ak,2M_k} \\ \vdots & \ddots & \vdots \\ p_{ak,M_k1} & \cdots & p_{ak,M_kM_k} \end{bmatrix} = \begin{bmatrix} \pi_{ak}^1 \\ \pi_{ak}^2 \\ \vdots \\ \pi_{ak}^{M_k} \end{bmatrix}, \tag{10}$$

where $P_{ak}^{M_k}$ denotes the steady-state probability matrix of state space $\mathbf{f}_{ak}$ with elements π_{ak}^i, $Q_{ak}^{M_k}$ denotes the corresponding state transition probability matrix with elements $p_{ak,mn}$. Especially, $p_{ak,mn}$ represents the state transition probability from state f_{ak}^m to state f_{ak}^n. Similar to $P_{ak}^{M_k}$, $P_{kb}^{N_k}$ can be obtained accordingly.

For finite-state Markov channel, both $f_{ak}(n)$ and $f_{kb}(n)$ are changing, where the joint steady-state probability of states f_{ak}^i and f_{kb}^j can be expressed as

$$\pi_{ij}^k = \pi_{ak}^i \pi_{kb}^j. \tag{11}$$

Based on (11), the average EST of R_k can be calculated as

$$\overline{T}^k = \sum_{j=1}^{N_k} \sum_{i=1}^{M_k} \pi_{ij}^k T_{ij}^k, \tag{12}$$

where T_{ij}^k denotes the EST of R_k when the broadcast channel and relay channel states are f_{ak}^i and f_{kb}^j, respectively.

B. Random and optimal selection schemes

Hereby, we investigate the achievable EST for RSS and OSS, respectively.

For the RSS, all the relays are selected randomly with equal probability, thus, the achievable EST of this scheme is asymptotic to average EST derived above, which is given by

$$\overline{T}_{rand} = \frac{1}{K} \sum_{k=1}^{K} \overline{T}^k. \tag{13}$$

For the OSS, the relay that leads to the maximal EST will be selected at each time slot based on the real-time CSI which is hard to realize in practical scenarios and the average EST of this scheme can be expressed as

$$\overline{T}_{opt} = \sum_{j_1=1}^{N_1} \cdots \sum_{j_K=1}^{N_K} \sum_{i_1=1}^{M_1} \cdots \sum_{i_K=1}^{M_K} \pi_{a1}^{i_1} \pi_{1b}^{j_1} \cdots \pi_{aK}^{i_K} \pi_{Kb}^{j_K} T_{max}, \tag{14}$$

where $T_{max} = \max\left(T_{i_1 j_1}^1, \cdots, T_{i_K j_K}^K\right)$.

Obviously, the EST achieved by OSS is much higher than that achieved by RSS. However, OSS is hard to realize in practical scenario due to the real time requirement of channel information feedback to Alice in wireless networks, where the CSI to select relay is usually outdated. To tackle this problem, we propose a learning-assisted relay selection methodology that is capable of properly selecting relay even with outdated CSI.

IV. Q-learning Assisted Relay Selection

In this section, we present our proposed Q-learning assisted relay selection scheme (QSS) to select secure relay with delayed CSI for finite-state Markov channel model, where $f_{ak}(n)$ and $f_{kb}(n)$ are outdated for relay selection. In particular, we will formulate the problem (8) as sequential decision process, which can be solved efficiently by our proposed learning-assisted scheme.

As a kind of reinforcement learning, Q-learning can realize sequential decision for Markov decision process (MDP) [15]. By interacting with the environment, the agent can learn the change of environment, and make decisions based on the experience obtained in the past rounds. Specifically, in each round, agent first observes the state of environment and executes an action selected from the action set $\mathcal{A}$ according to the state and the selection policy $Q(\mathcal{S}, \mathcal{A})$. Once the action a_n is executed, the corresponding reward $r(s_n, a_n)$ can be obtained, and the selection policy $Q(\mathcal{S}, \mathcal{A})$ can updated as follows

$$\begin{aligned} Q(s_n, a_n) \leftarrow (1-\alpha) Q(s_n, a_n) \\ + \alpha \left[r(s_n, a_n) + \beta \max_{a' \in \mathcal{A}} Q(s_{n+1}, a') \right], \end{aligned} \tag{15}$$

where s_n and a_n denote the observed state and the selected action at the n-th round, respectively, $\alpha \in (0, 1]$ is the learning rate, $\beta \in (0, 1)$ is the discount factor and $a' \in \mathcal{A}$ is the possible action that maximizes $Q(s_{n+1}, a')$.

It is evident that the trade-off between exploitation and exploration is crucial to the action selection process in Q-learning [16]. To this end, ε-greedy policy is employed to select the action obeying the following criterion

$$\Pr(a_n = a) = \begin{cases} 1-\varepsilon, & a = \operatorname{argmax}_{a \in \mathcal{A}} Q(s_n, a) \\ \dfrac{\varepsilon}{|\mathcal{A}| - 1}, & \text{o.w.} \end{cases}, \tag{16}$$

where $|\mathcal{A}|$ is the cardinality of set $\mathcal{A}$. Obviously, the ε-greedy policy is capable of selecting each action in $\mathcal{A}$ and obtaining the optimal action with the largest probability.

With the help of Q-learning, relay selection can be realized even $f_{ak}(n)$ and $f_{kb}(n)$ are outdated. The channel information of time slot $(n-1)$ is chosen as the state of time slot n, i.e. $s_n = [(f_{a1}(n-1), f_{1b}(n-1)), \cdots, (f_{aK}(n-1), f_{Kb}(n-1))]$. The available relays and the achievable EST of current time slot are chosen as the action set and reward, respectively, where $\mathcal{A} = [1, \cdots, K]$ and $r(s_n, a_n) = T^k$. In **Algorithm** 1,

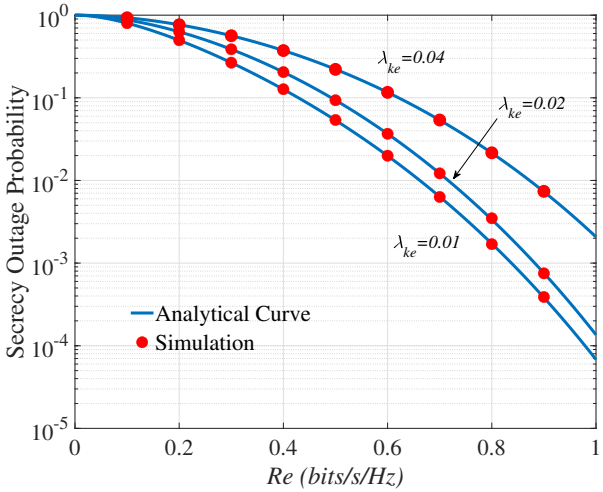

Fig. 2. P_s^k versus R_e for different λ_{ke} with λ_{ae}=0.03.

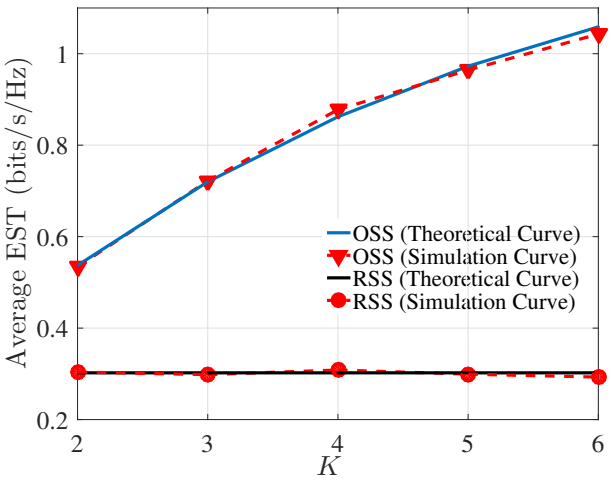

Fig. 3. Achievable EST of RSS and OSS with different K.

we demonstrate the selection of the secure relay according to the outdated CSI.

Algorithm 1 Q-learning assisted relay selection

Initialize $\alpha \in (0,1]$, $\beta \in (0,1)$, $\varepsilon > 0$, $Q(\mathcal{S},\mathcal{A}) = 0$.
for $n = 1, 2, \cdots, T$ **do**
  Observe state s_n.
  if rand()$< \varepsilon$ **then**
    Select all relays randomly.
  else
    Select the relay via $a_n = \arg\max_{a\in\mathcal{A}} Q(s_n, a)$.
  end if
  Calculate $r(s_n, a_n)$ via (7).
  Observe the next state s_{n+1}.
  Update $Q(s_n, a_n)$ via (15).
end for

V. Numerical Results

In this section, we present the simulation results to verify the accuracy of the theoretical analysis and the effectiveness of our proposed QSS. Specifically, we first show the accuracy of the secrecy outage probability. Then, we compare the achievable EST of the RSS and OSS. Finally, we compare the average EST obtained by different schemes.

In particular, we set $P_a = P_r = 10$ and $\sigma^2 = 1$, $\lambda_{ae} = 0.03$, and $\lambda_{1e} = \lambda_{2e} = \cdots = \lambda_{Ke} = 0.02$ following [17] throughout this section. As for the secrecy rate, we set $R_b = 2$ bits/s/Hz and $R_e = 0.5$ bits/s/Hz. Besides, we assume the sizes of the state spaces of the broadcasting channel and relay channels are 3, i.e. $M_k = N_k = 3 (k = 1, \cdots, K)$, where $\mathbf{f}_{ak} = [0.1, 1, 2]$ and $\mathbf{f}_{kb} = [1, 2, 3]$. Without loss of generality, we assume the transition probability $Q_{ak}^{M_k}$ and $Q_{kb}^{N_k}$ are the same for $k = 1, \cdots, K$, which can be expressed as

$$Q_{ak}^{M_k} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix},$$

$$Q_{kb}^{N_k} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0 & 0.9 \end{bmatrix}.$$

Furthermore, based on (10), the steady-state probability of each channels can be calculated as

$$P_{ak}^{M_k} = [0.4167, 0.25, 0.3333]^T,$$
$$P_{kb}^{N_k} = [0.3333, 0.1667, 0.5]^T.$$

In Fig.2, we verify the accuracy of the expression of the secrecy outage probability. Specifically, we plot secrecy outage probability for different λ_{ke} in this figure. It is obvious that the simulation points match with the analytical curve, which verifies the accuracy of our derivation. It also shows that P_s^k decreases as the increases of R_e, indicating that the secrecy performance can be improved by increasing the redundancy rate of the wiretap code.

In Fig.3, we plot the achievable EST of RSS and OSS with different K. It is evident that the theoretical curves, generated from (13) and (14), match the simulation points closely, thus validating the accuracy of our analysis. We also see that OSS achieves a higher EST than RSS, showing the importance of proper relay selection. Moreover, the EST obtained by OSS increases as the increase of K, while K has no influence to the EST obtained by RSS, this is because OSS can utilize the performance gain of multiple relays, while RSS cannot select the secure relay from multiple relays.

After that, we plot the curves of average EST for RSS and our proposed QSS with $K = 3$ in Fig. 4. Here, the learning parameters are $\varepsilon = 0.1$, $\alpha = 0.1$ and $\beta = 0.98$. It is apparent that our proposed QSS converges and achieves a higher EST

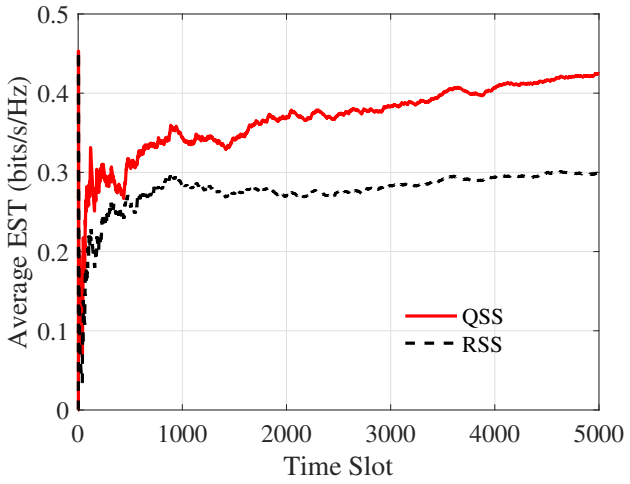

Fig. 4. Average EST for different schemes with $K = 3$.

compared to RSS, demonstrating the effectiveness of QSS when the CSI sent to Alice is outdated.

VI. Conclusion

In this paper, we proposed a learning-assisted secure relay selection methodology with outdated CSI for DF relay wireless networks, where the source selects a relay to forward the confidential information in the presence of a passive eavesdropper. Specifically, we firstly analyzed the achievable EST of RSS and OSS, and showed the importance of properly selecting relay in each time slot. Furthermore, Q-learning assisted scheme is proposed which could improve the secrecy performance even with outdated CSI.

Acknowledgement

This work was supported in part by the National Key R&D Program of China under Grant 2018YFB1801102, in part by National Natural Science Foundation of China (Grant No.61871253), and in part by Postdoctoral Science Foundation of China under Grant 2020M670332. *(Corresponding author: Zhaocheng Wang.)*

Appendix A
Derivation of (6)

Based on the definition of secrecy outage probability, we have

$$P_s^k = \Pr\left(2^{2R_e} - 1 < \frac{P_a f_{ae}(n)}{\sigma^2} + \frac{P_r f_{ke}(n)}{\sigma^2}\right) = 1 - \Pr\left(2^{2R_e} - 1 \geq \frac{P_a f_{ae}(n)}{\sigma^2} + \frac{P_r f_{ke}(n)}{\sigma^2}\right). \quad (17)$$

Define $P_a f_{ae}(n)/\sigma^2 + P_r f_{ke}(n)/\sigma^2$ as z^k. To further find the closed form expression of (17), we have to find the probability density function (PDF) of z^k first. For $f_{ae}(n) \sim exp\left(\frac{1}{\lambda_{ae}}\right)$, $f_{ke}(n) \sim exp\left(\frac{1}{\lambda_{ke}}\right)$ and $z^k = \frac{P_a f_{ae}(n)}{\sigma^2} + \frac{P_r f_{ke}(n)}{\sigma^2}$, where $x \sim exp(1/\lambda)$ denotes the random variable x following exponentially distribution with expectation λ, we have

$$p_{z^k}(z) = \int_0^z \frac{\sigma^2}{P_a\lambda_{ae}} e^{-\frac{\sigma^2}{P_a\lambda_{ae}}x} \frac{\sigma^2}{P_r\lambda_{ke}} e^{-\frac{\sigma^2}{P_r\lambda_{ke}}(z-x)} dx = \frac{\sigma^2}{P_a\lambda_{ae} - P_r\lambda_{ke}} \left(e^{-\frac{\sigma^2}{P_a\lambda_{ae}}z} - e^{-\frac{\sigma^2}{P_r\lambda_{ke}}z}\right). \quad (18)$$

Substituting (18) into (17), the closed form expression of secrecy outage probability can be derived as

$$P_s^k = 1 - \int_0^{2^{2R_e}-1} p_{z^k}(z)dz = \frac{\Theta^k}{\lambda_1}\left(e^{-\lambda_1(2^{2R_e}-1)}\right) - \frac{\Theta^k}{\lambda_2}\left(e^{-\lambda_2(2^{2R_e}-1)}\right). \quad (19)$$

The proof is completed.

References

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.* vol. 54, no. 8, pp. 1355-1387, Oct. 1975.
[2] J. Mo, M. Tao and Y. Liu, "Relay Placement for Physical Layer Security: A Secure Connection Perspective," *IEEE Trans. Commun. Lett.*, vol. 58, no. 3, pp. 878-881, Jun. 2012.
[3] D. W. K. Ng, E. S. Lo and R. Schober, "Secure Resource Allocation and Scheduling for OFDMA Decode-and-Forward Relay Networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3528-3540, Oct. 2011.
[4] Y. Zou, B. Champagne, W. Zhu and L. Hanzo, "Relay-Selection Improves the Security-Reliability Trade-Off in Cognitive Radio Systems," *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 215-228, Jan. 2015.
[5] W. Wang, K. C. Teh and K. H. Li, "Relay Selection for Secure Successive AF Relaying Networks With Untrusted Nodes," *IEEE Trans. Inf. Forensics Security.*, vol. 11, no. 11, pp. 2466-2476, Nov. 2016.
[6] D. S. Michalopoulos, H. A. Suraweera, G. K. Karagiannidis and R. Schober, "Amplify-and-Forward Relay Selection with Outdated Channel Estimates,"*IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1278-1290, May 2012.
[7] A. Tukmanov, S. Boussakta, Z. Ding and A. Jamalipour, "Outage Performance Analysis of Imperfect-CSI-Based Selection Cooperation in Random Networks," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2747-2757, Aug. 2014.
[8] Y. Su, L. Jiang and C. He, "Joint Relay Selection and Power Allocation for Full-Duplex DF Co-Operative Networks With Outdated CSI," IEEE Commun. Lett., vol. 20, no. 3, pp. 510-513, Mar. 2016.
[9] D. He, C. Liu, T. Q. S. Quek and H. Wang, "Transmit Antenna Selection in MIMO Wiretap Channels: A Machine Learning Approach,"*IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 634-637, Aug. 2018.
[10] D. He, C. Liu, H. Wang and T. Q. S. Quek, "Learning-Based Wireless Powered Secure Transmission," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 600-603, Apr. 2019.
[11] Z. Miao and Y. Wang, "Physical-Layer-Security-Oriented Frequency Allocation in Ultra-Dense-Networks Based on Location Informations," *IEEE Access*, vol. 7, pp. 90190-90205, 2019.
[12] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels,"*IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688-1692, Nov. 1999.
[13] J. Yao, S. Feng, X. Zhou and Y. Liu, "Secure Routing in Multihop Wireless Ad-Hoc Networks With Decode-and-Forward Relaying," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 753-764, Feb. 2016.
[14] S. Yan, N. Yang, G. Geraci, R. Malaney and J. Yuan, "Optimization of Code Rates in SISOME Wiretap Channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6377-6388, Nov. 2015.
[15] M. A. Jadoon and S. Kim, "Relay selection Algorithm for wireless cooperative networks: a learning-based approach," *IET Commun.*, vol. 11, no. 7, pp. 1061-1066, May 2017.
[16] R.S. Sutton, A.G. Barto, *Reinforcement learning: an introduction*. Cambridge University Press, Cambridge, 2011.
[17] L. Yang, J. Chen, H. Jiang, S. A. Vorobyov and H. Zhang, "Optimal Relay Selection for Secure Cooperative Communications With an Adaptive Eavesdropper," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 26-42, Jan. 2017.