Enhancing physical-layer security via big-data-aided hybrid relay selection

Hongliang He^{1,2}, Pinyi Ren^{1,2*}, Qinghe Du^{1,2}, Li Sun^{1,2}, Yichen Wang^{1,2}

1. School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

2. Shaanxi Smart Networks and Ubiquitous Access Research Center, Xi'an 710049, China

* Corresponding author, email: pyren@mail.xjtu.edu.cn

Abstract: The explosive growth in data traffic presents new challenges to the new generation of wireless communication systems, such as computing capabilities, spectrum efficiency and security. In this paper, we use the network structure, which is adaptable for the big data traffic, to improve the security of wireless networks. Specifically, a big-data aided hybrid relay selection scheme is designed and analyzed to enhance physical layer security. First, considering the ideal situation that an eavesdropper's CSI (Channel State Information) is known to the legal nodes, we propose an optimal hybrid relay selection scheme consisting of the optimal mode selection scheme and the optimal relay selection scheme. In this case, we analyze the upper bound of an eavesdropper's capacity in FD (Full-Duplex) mode and the secrecy outage probabilities of the optimal HD (Half-Duplex), FD, and hybrid relay selection schemes. Through the analysis of data, it is clear that the mode selection is decided by the self-interference of the FD technique. However, the instantaneous CSI of an eavesdropper is difficult to obtain due to the passive characteristic of eavesdroppers in practice. Therefore, a more practical hybrid relay selection scheme with only the channel distribution information of an eavesdropper is further studied, where a weighting factor is employed to guarantee that the hybrid mode is no worse than either the FD mode or HD mode when the self-interference grows. Finally, the simulation results show the improved security of our proposed scheme.

Keywords: Big data, hybrid relay selection, secrecy outage probability, eavesdropper, physical layer security

Citation: H. L. He, P. Y. Ren, Q. H. Du, et al. Enhancing physical-layer security via big-data-aided hybrid relay selection [J]. Journal of communications and information networks, 2017, 2(1): 97-110.

1 Introduction

Security and privacy have become serious concerns in the big-data era with the explosive growth of data traffic. The traditional schemes for improving information security focus on cryptographic techniques using higher-layer protocols. However, physical layer security achieves secrecy by exploiting the inherent randomness of wireless channels without the help of an encryption key. Recently, relay selection^[1-8] has been considered as one of the most promising physi-

cal layer security schemes^[9-13] to enhance security by protecting private information from eavesdropping in wireless communications systems. A typical assumption in existing research is that the relay applies HD mode with two orthogonal time slots or equivalently two orthogonal frequency bands allocated for the respective reception and transmission. The FD mode overcomes this problem by simultaneously transmitting and receiving on the same frequency. However, its implementation is limited by the self-interference caused by signal leakage between the output and in-

Manuscript received Dec. 15, 2016; accepted Feb. 08, 2017

This work is supported by the National Natural Science Foundation of China (NSFC) (No. 61431011), the Fundamental Research Funds for the Central Universities.

put. Thus, a hybrid FD/HD technique with opportunistic switching to the optimal mode has been developed to maximize spectral efficiency. However, it is not effortless to combine the hybrid FD/HD technique and relay selection technique for use with intensely high system traffic as well as the rapidly growing data. With a large number of relays, calculating and comparing the secrecy capacity in different relays and different modes imposes an enormous burden on BS (Base Station). Therefore, it is necessary to employ an efficient wireless architecture as well as signal processing method to adapt to the growth in data traffic. Fortunately, with the development of big data^[14-16], the cloud-based cellular architecture is likely to replace the BS-centric cellular architecture in 5G wireless networks. This will prompt the achievement of cooperative communication with intensely high system traffic and data. Thus, a big-data-aided hybrid relay selection scheme is proposed in this paper to improve the security of wireless transmission.

HD relay selection^[1-6] has been studied extensively to enhance the physical layer security. In general, one or several relays are selected to improve the secrecy rate. For example, a relay is selected to maximize the secrecy rate in Ref. [5]; Two relays are selected in Ref. [4], where one relay is employed for forwarding private messages and the other is utilized to emit artificial noise. In Ref. [6], diversity orders and the probabilities of positive secrecy capacity are analyzed under various relay selection schemes. Although FD relay selection^[7,8] has been researched in some studies, few of them consider the security constraint. Some research^[17-19] has analyzed the performance of FD relay on physical layer security, but only one relay was considered instead of FD relay selection. Specifically, in Ref. [17], a composite fading channel model containing Nakagami- m fading as well as log-normal shadowing was employed to study the performance of secure FD relaying. In Ref. [18], the security performance of FD and HD techniques was compared. Secure communication in multi-hop FD relaying systems was considered in Ref. [19]. However, it was assumed in Ref. [19] that

the self-interference can be perfectly canceled, which might be too optimistic. In Ref. [20], SOPs (Secrecy Outage Probabilities) with FD and HD relay were analyzed, and a jamming scheme was proposed to further improve the secrecy performance. A similar method was considered in Ref. [21]. A hybrid FD/HD technique was proposed in Refs. [22,23], but the security constraint was not considered.

In contrast with the HD or FD relay selection scheme, this paper focuses on the design and analysis of a hybrid HD/FD relay selection scheme with physical layer security from the perspective of big-data. The upper bound of an eavesdropper's capacity in FD mode is first studied. Based on this, an optimal mode selection scheme without relay selection is proposed under the assumption that an eavesdropper's CSI is known. Then the SOPs of the optimal HD, FD, and hybrid relay selection schemes are analyzed. However, the instantaneous CSI of an eavesdropper is difficult to obtain due to the passive characteristic of eavesdroppers in practice. Therefore, a more practical hybrid relay selection scheme with only the channel distribution information of an eavesdropper is studied in this paper, in which a weighting factor is employed to guarantee that the hybrid mode is no worse than either the FD mode or HD mode.

The remainder of the paper is organized as follows. Section 2 presents the system model. Section 3 describes the optimal hybrid relay selection scheme. Section 4 provides SOP (Secrecy Outage Probability) of the proposed scheme. Section 5 analyzes the hybrid relay selection scheme without the eavesdropper's instantaneous CSI (ECSI). Section 6 shows the simulation results used to evaluate the performance of the proposed scheme. The paper is concluded in section 7.

2 System model

Consider that a great number of users with common security interest compose a relay network, as shown in Fig. 1. The source S intends to transmit private messages to the legal destination D by selecting a relay R_k ($k \in \mathbb{N}$, $\mathbb{N} = \{1, 2, \dots, N\}$) from N trusted

relays, but the private messages are intercepted by an eavesdropper E .

Figure 1 System model

The selection of a relay is the result of a calculation by a data fusion center, i.e., a cloud-centric processor. All relays employing the decode-and-forwarding protocol are equipped with two antennas, one receiving antenna and one transmitting antenna, and the other users have only one antenna. When a relay operates in FD mode, it suffers from self-interference. Although the ideal FD relay cannot be achieved at present, some schemes can be used to mitigate self-interference, such as antenna isolation, spatial precoding and time cancellation in Refs. [24,25]. There is no direct communication link between the source S and the legitimate destination D because of the long distance and high attenuation of signals, but the eavesdropper can intercept the information from both the source and the relay. Denote the channel between node $m \in \{s, r\}$ and node $n \in \{r, d, e\}$ as h_{mn} . All channels are assumed to experience quasi-static Rayleigh fading and are independent of each other. The channel coefficient h_{mn} remains constant over one block and varies independently from one block to another, which is modeled as a circularly symmetric complex Gaussian distribution with zero mean and variance σ_{mn}^2 . The channels from the source to the relay, from the relays to destination, and from the relays to the eavesdropper are considered independent and identically distributed (i.i.d.) with variance σ_{sr}^2 , σ_{rd}^2 , and σ_{re}^2 , respectively. The self-interference can be mitigated and its resid-

ual can also be considered as following a circularly symmetric complex Gaussian distribution (See Refs. [7,25,26] and the references therein). Thus, the notation $h_{r_k r_k}$ is employed to denote the k -th relay's self-interference channel with zero mean and variance σ_{rr}^2 . Without loss of generality, all noise is assumed to be complex and independent AWGN (Additive White Gaussian Noise) with zero and unit variance. In addition, the transmission power of the source and the relay are fixed with p_s and p_r respectively. Throughout this paper, $\text{lb}(\cdot)$ denotes base-2 logarithm, $[x]^+ = \max(0, x)$, $(\cdot)^\dagger$ denotes conjugate transpose, $(\cdot)^{-1}$ denotes inverse, $(\cdot)^T$ represents transpose, and $\text{E}\{\cdot\}$ is the expectation operation. The channel SNR (Signal-to-Noise Ratio) is $\gamma_{mn} = p_m |h_{mn}|^2$ obeying an exponential distribution with parameter $\lambda_{mn} = 1/(p_m \sigma_{mn}^2)$, and its probability density function (p.d.f.) can be described as

$$f_{m,n}(x) = \lambda_{mn} e^{-\lambda_{mn} x}, \quad x > 0. \quad (1)$$

2.1 FD mode

The frequency is employed to receive and forward simultaneously, so the transmission is degraded by self-interference. At time slot i , the source transmits $x(i)$ with $\text{E}\{|x(i)|^2\} = p_s$. Relay receives $r(i)$ and transmits $z(i)$ simultaneously with $\text{E}\{|z(i)|^2\} = p_r$. Both $x(i)$ and $z(i)$ are complex Gaussian variables. Thus, the received signals at relay R_k , destination D and eavesdropper E are given by

$$r(i) = h_{sr_k} x(i) + h_{r_k r_k} z(i) + n_{r_k}(i), \quad (2)$$

$$y(i) = h_{r_k d} z(i) + n_d(i), \quad (3)$$

$$y_e(i) = h_{se} x(i) + h_{r_k e} z(i) + n_e(i), \quad (4)$$

respectively, where $n_r(i)$, $n_d(i)$, and $n_e(i)$ denote the noise. Owing to the power constraint and processing delay $\tau \geq 1$, the relay R_k forwards $x(i - \tau)$ in slot i , and the forwarding signal can be expressed as

$$z(i) = \sqrt{\frac{p_r}{p_s}} x(i - \tau). \quad (5)$$

We assume that $\tau = 1$ is large enough to guarantee the uncorrelation of the received signals, so $\tau = 1$ is employed in this paper hereafter.

Consider that the length of a block is $L + \tau$, i.e., the source transmits L successive codewords and all channel gains remain constant over the block. If the eavesdropper employs joint decoding, its received signal can be written in vector form as

$$\mathbf{y}_e = \mathbf{H}_e \mathbf{x} + \mathbf{n}_e, \quad (6)$$

where $\mathbf{x} = (x(1), \dots, x(L))^T$, $\mathbf{y}_e = (y_e(1), \dots, y_e(L + \tau))^T$, $\mathbf{n}_e = (n_e(1), \dots, n_e(L + \tau))^T$, and

$$\mathbf{H}_e = h_{se} \begin{bmatrix} \mathbf{I}_L \\ \mathbf{0}_{\tau \times L} \end{bmatrix} + \sqrt{\frac{p_r}{p_s}} h_{rke} \begin{bmatrix} \mathbf{0}_{\tau \times L} \\ \mathbf{I}_L \end{bmatrix}, \quad (7)$$

where $\mathbf{0}_{\tau \times L}$ denotes a $\tau \times L$ zero matrix, and $\mathbf{I}_L$ denotes an $L \times L$ unit matrix.

2.2 HD mode

Two orthogonal time slots are required to transmit the private signal. The source transmits $x(2i)$ during the even slots, and the relay receives $r(2i)$ during the even slots and forwards $z(2i+1)$ during the odd slots. Thus, the received signals at relay R_k , destination D and eavesdropper E are given by

$$r(2i) = h_{srk} x(2i) + n_{rk}(2i), \quad (8)$$

$$y(2i+1) = h_{rkd} z(2i+1) + n_d(2i+1), \quad (9)$$

$$y_e(2i) = h_{se} x(2i) + n_e(2i), \quad (10)$$

$$y_e(2i+1) = h_{rke} z(2i+1) + n_e(2i+1), \quad (11)$$

respectively. Owing to the power constraint, we have $z(2i+1) = \sqrt{p_r/p_s} x(2i)$.

3 Hybrid HD/FD scheme

In this section, the HD, FD, and optimal hybrid relay selection schemes are studied in theory.

3.1 Security capacity of the FD/HD mode

According to the received signal discussed above, the capacity of the legitimate destination can be described as in Ref. [22]

$$C_{d,k}^{\text{FD}} = \text{lb} \left(1 + \min \left\{ \frac{\gamma_{srk}}{\gamma_{rkrk} + 1}, \gamma_{rkd} \right\} \right), \quad (12)$$

where $\gamma_{srk} = p_s |h_{srk}|^2$, $\gamma_{rkrk} = p_r |h_{rkrk}|^2$, and $\gamma_{rkd} = p_r |h_{rkd}|^2$.

From Eq. (6), we obtain the capacity of the eavesdropper as

$$\begin{aligned} C_{e,k}^{\text{FD}} &= \frac{1}{L} \text{lb det} (\mathbf{I}_L + p_s \mathbf{H}_e^\dagger \mathbf{H}_e) \\ &= \frac{1}{L} \text{lb} \prod_{i=1}^L (1 + p_s \theta_i), \end{aligned} \quad (13)$$

where θ_i ($i = 1, \dots, L$) denotes the eigenvalues of $\mathbf{H}_e^\dagger \mathbf{H}_e$. It is important to find that $\mathbf{H}_e^\dagger \mathbf{H}_e = \mu_e \mathbf{I}_L + \nu_e \mathbf{A}_{L,e} + \nu_e^* \mathbf{B}_{L,e}$ with $\mu_e = |h_{se}|^2 + (p_r/p_s) |h_{rke}|^2$ and $\nu_e = \sqrt{p_r/p_s} h_{se}^* h_{rke}$. Here, $\mathbf{A}_{L,e}$ ($\mathbf{B}_{L,e}$) denotes an $L \times L$ backward (forward) shift matrix, with ones only on the first subdiagonal (superdiagonal) and zeros elsewhere. Matrix $\mathbf{H}_e^\dagger \mathbf{H}_e$ is a tridiagonal Toeplitz matrix when $\tau = 1$, whose eigenvalues can be written as in Ref. [27].

$$\theta_i = \mu_e + 2|\nu_e| \cos \frac{i\pi}{L+1}, i \in \{1, \dots, L\}. \quad (14)$$

Thus, the capacity of the eavesdropper can be further described as

$$\begin{aligned} C_{e,k}^{\text{FD}} &= \frac{1}{L} \text{lb} \prod_{i=1}^L \left(1 + p_s \mu_e + 2p_s |\nu_e| \cos \frac{i\pi}{L+1} \right) \\ &= \text{lb}(1 + p_s \alpha) + \frac{1}{L} \sum_{i=1}^L \text{lb} \left(1 + \frac{2p_s |\nu_e| \cos \frac{i\pi}{L+1}}{1 + p_s \alpha} \right). \end{aligned} \quad (15)$$

It is clear that $\mu_e \geq 2|\nu_e|$ according to the arithmetic-geometric mean inequality, meaning that $2p_s |\nu_e| \cos \frac{i\pi}{L+1} \leq 2p_s |\nu_e| \leq p_s \mu_e \leq 1 + p_s \mu_e$ and $\left| 2p_s |\nu_e| \cos \frac{i\pi}{L+1} / (1 + p_s \mu_e) \right| < 1$ are always satisfied. It is worth noting that $f(x) = \ln(1+x) - x$ is a monotone decreasing function with $f(0) = 0$, which is why inequality $\ln(1+x) < x$ is always satisfied. Combining it with the limit $\lim_{x \rightarrow 0} \ln(1+x)/x = 1$, we discover that the upper bound of $\ln(1+x)$ is x , and they are really close with small x (e.g., $|x| < 1$). Therefore, the upper bound of $C_{e,k}^{\text{FD}}$ is acquired as

$$C_{e,k}^{\text{FD,UB}}$$

$$\begin{aligned}
&= \text{lb}(1 + p_s \mu_e) + \frac{1}{L} \sum_{i=1}^L \frac{2p_s |\nu_e| \cos \frac{i\pi}{L+1}}{(1 + p_s \mu_e) \ln 2} \\
&= \text{lb}(1 + p_s \mu_e) + \frac{2p_s |\nu_e|}{\ln 2 \cdot L(1 + p_s \mu_e)} \sum_{i=1}^L \cos \frac{i\pi}{L+1},
\end{aligned} \quad (16)$$

which shows the inequality $C_{e,k}^{\text{FD}} \leq C_{e,k}^{\text{FD-UB}}$. It is obvious that $\sum_{i=1}^L \cos(i\pi/(L+1)) = 0$ because the equality $\cos \alpha + \cos(\pi - \alpha) = 0$ is always true, which demonstrates that the second term of Eq. (16) is zero. The upper bound of C_e^{FD} is then given by

$$C_{e,k}^{\text{FD-UB}} = \text{lb}(1 + p_s \mu_e) = \text{lb}(1 + \gamma_{se} + \gamma_{rke}), \quad (17)$$

where $\gamma_{se} = p_s |h_{se}|^2$, $\gamma_{rke} = p_r |h_{rke}|^2$.

Thus, the secrecy capacity can be written as

$$\begin{aligned}
C_{s,k}^{\text{FD}} &= [C_{d,k}^{\text{FD}} - C_{e,k}^{\text{FD}}]^+ \\
&= \left[\text{lb} \frac{1 + \min \left\{ \frac{\gamma_{srk}}{\gamma_{rkr} + 1}, \gamma_{rkd} \right\}}{\left(\det \{ \mathbf{I}_L + p_s \mathbf{H}_e^\dagger \mathbf{H}_e \} \right)^{\frac{1}{L}}} \right]^+,
\end{aligned} \quad (18)$$

and the lower bound of the security capacity can be written as

$$\begin{aligned}
C_{s,k}^{\text{FD-LB}} &= [C_d^{\text{FD}} - C_e^{\text{FD-UB}}]^+ \\
&= \left[\text{lb} \left(\frac{1 + \min \left\{ \frac{\gamma_{srk}}{\gamma_{rkr} + 1}, \gamma_{rkd} \right\}}{1 + \gamma_{se} + \gamma_{rke}} \right) \right]^+ \\
&= [\text{lb}(\gamma_{\text{FD}})]^+.
\end{aligned} \quad (19)$$

It is obvious that inequality $C_{s,k}^{\text{FD}} \geq C_{s,k}^{\text{FD-LB}}$ is true. Since Eq. (18) is difficult to analyze, in this paper we henceforth employ the lower bound of the security capacity instead of the actual secrecy capacity.

In HD mode, we assume that the eavesdropper adopts MRC (Maximal Ratio Combining). Therefore, the capacity^[10] of the legitimate destination and eavesdropper are given by

$$C_{d,k}^{\text{HD}} = \frac{1}{2} \text{lb} \left(1 + \min \{ \gamma_{srk}, \gamma_{rkd} \} \right), \quad (20)$$

$$C_{e,k}^{\text{HD}} = \frac{1}{2} \text{lb} (1 + \gamma_{se} + \gamma_{rke}), \quad (21)$$

Then we obtain the secrecy capacity of the HD mode as

$$C_{s,k}^{\text{HD}} = [C_{d,k}^{\text{HD}} - C_{e,k}^{\text{HD}}]^+$$

$$\begin{aligned}
&= \left[\frac{1}{2} \text{lb} \left(\frac{1 + \min \{ \gamma_{srk}, \gamma_{rkd} \}}{1 + \gamma_{se} + \gamma_{rke}} \right) \right]^+ \\
&= \left[\text{lb} \sqrt{\frac{1 + \min \{ \gamma_{srk}, \gamma_{rkd} \}}{1 + \gamma_{se} + \gamma_{rke}}} \right]^+ \\
&= [\text{lb}(\gamma_{\text{HD}})]^+.
\end{aligned} \quad (22)$$

3.2 Hybrid relaying mode

In this subsection, the hybrid relaying mode without a relay section is considered. Since the FD mode is not always better than the HD mode due to the effect of self-interference, the hybrid relaying scheme enables relay switching between FD mode and HD mode to achieve better performance. If the secrecy capacity of the HD mode is superior to the FD mode, the relay employs the HD mode; otherwise, it employs the FD mode. This process can be described as

$$\begin{cases} \text{HDmode, if } C_{s,k}^{\text{HD}} > C_{s,k}^{\text{FD}}, \\ \text{FDmode, if } C_{s,k}^{\text{FD}} \geq C_{s,k}^{\text{HD}}. \end{cases}$$

Thus, the secrecy capacity of the hybrid relaying mode can be described as

$$C_{s,k}^{\text{Hybrid}} = \max \{ C_{s,k}^{\text{HD}}, C_{s,k}^{\text{FD}} \}. \quad (23)$$

Since the secrecy capacity of the FD mode is always larger than that of the HD mode when the self-interference is zero, the choice of operating mode is decided by the intensity of self-interference. Compare Eq. (18) with Eq. (21), there are three cases obtained as follows.

1) Case 1: $\gamma_{rkd} < \gamma_{srk}/(\gamma_{rkr} + 1) \Rightarrow 0 < \gamma_{rkr} < \gamma_{srk}/(\gamma_{rkd}) - 1$

In this case, the capacity of the FD mode is always larger than that of the HD mode. Because inequality $\gamma_{srk} > \gamma_{rkd}$ can be deduced, the variable $\gamma_{\text{FD}} = (1 + \gamma_{rkd})/(1 + \gamma_{se} + \gamma_{rke})$ is always larger than $\gamma_{\text{HD}} = \sqrt{(1 + \gamma_{rkd})/(1 + \gamma_{se} + \gamma_{rke})}$ when secrecy transmission is ensured, and the capacity of the FD mode is twice as large as that of the HD mode.

2) Case 2: $\gamma_{srk}/(\gamma_{rkr} + 1) \leq \gamma_{rkd} < \gamma_{srk}$

The values of $\gamma_{\text{FD}} = (1 + \gamma_{sr_k}/(\gamma_{r_k r_k} + 1))/(1 + \gamma_{se} + \gamma_{r_k e})$ and $\gamma_{\text{HD}} = \sqrt{(1 + \gamma_{r_k d})/(1 + \gamma_{se} + \gamma_{r_k e})}$ are acquired in this case. When the self-interference satisfies

$$\frac{\gamma_{sr_k}}{\gamma_{r_k d}} - 1 \leq \gamma_{r_k r_k} < \frac{\gamma_{sr_k}}{\sqrt{(1 + \gamma_{r_k d})(1 + \gamma_{se} + \gamma_{r_k e})} - 1} - 1,$$

we have $\gamma_{\text{FD}} > \gamma_{\text{HD}}$ and the FD mode is selected, i.e., $C_{s,k}^{\text{FD}} \geq C_{s,k}^{\text{HD}}$; When the self-interference satisfies

$$\gamma_{r_k r_k} \geq \frac{\gamma_{sr_k}}{\sqrt{(1 + \gamma_{r_k d})(1 + \gamma_{se} + \gamma_{r_k e})} - 1} - 1, \quad (24)$$

we have $\gamma_{\text{FD}} < \gamma_{\text{HD}}$, and the HD mode is selected.

3) Case 3: $\gamma_{sr_k} < \gamma_{r_k d}$

The values of $\gamma_{\text{FD}} = (1 + \gamma_{sr_k}/(\gamma_{r_k r_k} + 1))/(1 + \gamma_{se} + \gamma_{r_k e})$ and $\gamma_{\text{HD}} = \sqrt{(1 + \gamma_{sr_k})/(1 + \gamma_{se} + \gamma_{r_k e})}$ are obtained. Furthermore, when $\gamma_{\text{FD}} > \gamma_{\text{HD}}$, we have

$$\frac{\gamma_{sr_k}}{\gamma_{r_k d}} - 1 < \gamma_{r_k r_k} < \frac{\gamma_{sr_k}}{\sqrt{(1 + \gamma_{sr_k})(1 + \gamma_{se} + \gamma_{r_k e})} - 1} - 1,$$

and the FD mode is selected. Otherwise, when the inequation

$$\gamma_{r_k r_k} > \frac{\gamma_{sr_k}}{\sqrt{(1 + \gamma_{sr_k})(1 + \gamma_{se} + \gamma_{r_k e})} - 1} - 1,$$

is satisfied, the HD mode is selected.

In addition, it is clear that if $\gamma_{\text{FD}} > 1$ is true, inequality $\gamma_{\text{HD}} > 1$ is always true in all cases, meaning a positive secrecy capacity for the HD mode when the secrecy capacity of the FD mode is larger than zero.

3.3 Relay selection

The FD relay selection scheme is based on the secrecy capacity of the FD mode. The relay achieving the maximum secrecy capacity is chosen, which can be described as

$$k_f^* = \arg \max_k C_{s,k}^{\text{FD}}. \quad (25)$$

Similarly, the traditional HD relay selection scheme is given by

$$k_h^* = \arg \max_k C_{s,k}^{\text{HD}}. \quad (26)$$

The hybrid relay selection scheme providing a better performance than both the FD mode and HD mode is then given by

$$k^* = \arg \max_k \max\{C_{s,k}^{\text{HD}}, C_{s,k}^{\text{FD}}\}, \quad (27)$$

where $k \in \{1, \dots, N\}$.

4 SOP with ECSI

The SOP of the proposed scheme is analyzed in theory in this section, and is defined as the probability that the secrecy capacity is less than a target secrecy rate R_s .

4.1 SOP of the optimal FD relay selection

Since it is hard to analyze Eq. (18), its lower bound Eq. (19) is analyzed instead. For arbitrary relay R_k , the SOP without relay selection in FD mode is given by

$$\begin{aligned} & \Pr(C_{s,k}^{\text{FD}} < R_s) \\ &= \Pr \left\{ \text{lb} \left(\frac{1 + \min \left\{ \frac{\gamma_{sr_k}}{\gamma_{r_k r_k} + 1}, \gamma_{r_k d} \right\}}{1 + \gamma_{se} + \gamma_{r_k e}} \right) < R_s \right\}. \end{aligned} \quad (28)$$

Because the channels $h_{r_k r_k}$ are i.i.d., the subscript k is omitted during the calculation for convenience. We start by analyzing the cumulative density function (CDF) of $X_1 = \gamma_{sr}/(\gamma_{rr} + 1)$, which can be described as

$$\begin{aligned} F_{X_1}(x) &= \int_0^\infty F_{\gamma_{sr}}(x(y+1)) f_{\gamma_{rr}}(y) dy \\ &= 1 - \frac{e^{-\lambda_{sr}x}}{1 + \frac{\lambda_{sr}}{\lambda_{rr}}x}. \end{aligned} \quad (29)$$

Then the CDF of $Y_1 = \min\{\gamma_{sr}/(\gamma_{rr} + 1), \gamma_{rd}\}$ is given by

$$\begin{aligned} F_{Y_1}(x) &= F_{X_1}(x) + F_{\gamma_{rd}}(x) - F_{X_1}(x)F_{\gamma_{rd}}(x) \\ &= 1 - \frac{e^{-(\lambda_{sr} + \lambda_{rd})x}}{1 + \frac{\lambda_{sr}}{\lambda_{rr}}x}. \end{aligned} \quad (30)$$

Using the convolution formula, we obtain the p.d.f. of $Z_1 = \gamma_{se} + \gamma_{re}$ as

$$f_{Z_1}(z) = \frac{\lambda_{re}\lambda_{se}(e^{-\lambda_{se}z} - e^{-\lambda_{re}z})}{\lambda_{re} - \lambda_{se}}. \quad (31)$$

The CDF of $X_2 = \text{lb}\left((1+Y_1)/(1+Z_1)\right)$ is then given by

$$\begin{aligned} F_{X_2}(x) &= 1 - \frac{ma(x) \cdot e^{\frac{(\lambda_{se}+b(x))c(x)}{d(x)}}}{d(x)} \text{E}_1\left(\frac{(\lambda_{se}+b(x))c(x)}{d(x)}\right) \\ &\quad + \frac{ma(x) \cdot e^{\frac{(\lambda_{re}+b(x))c(x)}{d(x)}}}{d(x)} \text{E}_1\left(\frac{(\lambda_{re}+b(x))c(x)}{d(x)}\right), \end{aligned}$$

where $\text{E}_1(x) = \int_x^\infty e^{-t}t^{-1}dt$ denotes the exponential integral function, $m = \lambda_{re}\lambda_{se}/(\lambda_{re} - \lambda_{se})$, $a(x) = \exp\{-(\lambda_{sr} + \lambda_{rd})2^x + \lambda_{sr} + \lambda_{rd}\}$, $b(x) = (\lambda_{sr} + \lambda_{rd})2^x$, $c(x) = 1 + (\lambda_{sr}/\lambda_{rr})2^x - \lambda_{sr}/\lambda_{rr}$, $d(x) = 2^x \cdot \lambda_{sr}/\lambda_{rr}$. This process of calculation is provided in Appendix A.

Now, it is obvious that the SOP shown in Eq. (28) can be rewritten as

$$\begin{aligned} \Pr(C_{s,k}^{\text{FD}} < R_s) &= \Pr\left\{\text{lb}\left(\frac{1+Y_1}{1+Z_1}\right) < R_s\right\} \\ &= \Pr\{X_2 < R_s\} = F_{X_2}(R_s). \end{aligned} \quad (32)$$

According to the results of the order statistics random variable^[28], the SOP of the optimal FD relay selection can be acquired as

$$\begin{aligned} P_{\text{out}}^{\text{FDRS}} &= \Pr(C_{s,k^*}^{\text{FD}} < R_s) \\ &= \Pr(X_2^* < R_s) \\ &= \left(F_{X_2}(R_s)\right)^N, \end{aligned} \quad (33)$$

where C_{s,k^*}^{FD} denotes the secrecy capacity by selecting the optimal FD relay k^* according to the criterion in Eq. (25). Substituting Eq. (32) into Eq. (33), we obtain the following result:

$$\begin{aligned} P_{\text{out}}^{\text{FD}} &= \left(1 - \frac{ma(R_s) \cdot e^{\frac{(\lambda_{se}+b(R_s))c(R_s)}{d(R_s)}}}{d(R_s)} \cdot \text{E}_1\left(\frac{(\lambda_{se}+b(R_s))c(R_s)}{d(R_s)}\right) \right. \\ &\quad \left. + \frac{ma(R_s) \cdot e^{\frac{(\lambda_{re}+b(R_s))c(R_s)}{d(R_s)}}}{d(R_s)} \cdot \text{E}_1\left(\frac{(\lambda_{re}+b(R_s))c(R_s)}{d(R_s)}\right) \right)^N. \end{aligned}$$

$$\cdot \text{E}_1\left(\frac{(\lambda_{re}+b(R_s))c(R_s)}{d(R_s)}\right)^N. \quad (34)$$

4.2 SOP of the optimal HD relay selection

According to Eq. (22), we obtain the SOP without relay selection in HD mode as

$$\begin{aligned} \Pr(C_{s,k}^{\text{FD}} < R_s) &= \Pr\left\{\frac{1}{2}\text{lb}\left(\frac{1 + \min\{\gamma_{srk}, \gamma_{rkd}\}}{1 + \gamma_{se} + \gamma_{rke}}\right) < R_s\right\}. \end{aligned} \quad (35)$$

Denoting $Y_2 = \min\{\gamma_{sr}, \gamma_{rd}\}$, we obtain the CDF of Y_2 as

$$\begin{aligned} F_{Y_2}(x) &= F_{\gamma_{sr}}(x) + F_{\gamma_{rd}}(x) - F_{\gamma_{sr}}(x)F_{\gamma_{rd}}(x) \\ &= 1 - e^{-(\lambda_{sr}+\lambda_{rd})x}, \end{aligned} \quad (36)$$

Because the p.d.f. of $Z_1 = \gamma_{se} + \gamma_{re}$ is shown in Eq. (31), the CDF of $X_3 = (1/2) \cdot \text{lb}\left((1+Y_2)/(1+Z_1)\right)$ is then given by

$$\begin{aligned} F_{X_3}(x) &= 1 - m \cdot q(x) \cdot \left(\frac{1}{\lambda_{se} + p(x)} - \frac{1}{\lambda_{re} + p(x)}\right), \end{aligned} \quad (37)$$

where $m = \lambda_{re}\lambda_{se}/(\lambda_{re} - \lambda_{se})$, $p(x) = (\lambda_{sr} + \lambda_{rd}) \cdot 2^{2x}$, and $q(x) = \exp\{-(\lambda_{sr} + \lambda_{rd}) \cdot (2^{2x} - 1)\}$. The calculation is provided in Appendix B. Similar to the FD relay selection scheme, the SOP of the optimal HD relay selection is given by

$$\begin{aligned} P_{\text{out}}^{\text{HDRS}} &= \Pr(C_{s,k^*}^{\text{HD}} < R_s) = \Pr(X_3^* < R_s) \\ &= \left(F_{X_3}(R_s)\right)^N \\ &= \left(1 - \frac{\lambda_{re}\lambda_{se} \cdot q(R_s)}{(\lambda_{se} + p(R_s))(\lambda_{re} + p(R_s))}\right)^N. \end{aligned} \quad (38)$$

4.3 Big-data analysis of hybrid relay selection

According to the secrecy capacity of the hybrid relay scheme, the SOP of the optimal hybrid relaying selection can be described as

$$P_{\text{out}}^{\text{Hybrid}} = \left(\Pr\{C_{s,k}^{\text{Hybrid}} < R_s\}\right)^N$$

$$= \left(\Pr \left\{ \max \{ C_{s,k}^{\text{HD}}, C_{s,k}^{\text{FD}} \} < R_s \right\} \right)^N. \quad (39)$$

It is worth noting that the SOP of the optimal hybrid relay selection scheme is not equal to the product of the SOP of the optimal FD relay selection scheme and the SOP of the optimal HD relay selection scheme, because $C_{s,k}^{\text{HD}}$ and $C_{s,k}^{\text{FD}}$ are not independent. However, since the secrecy capacity of the hybrid scheme is larger than that of the HD and FD scheme, the SOP of the hybrid scheme is less than that of the HD and FD scheme.

It is well-known that $\lim_{N \rightarrow \infty} a^N = 0$ is true if we have $0 < a < 1$. Since the SOPs of the FD, HD, and the hybrid relay selection schemes in Eq. (32), Eq. (37) and Eq. (39) are all less than one, the SOPs will tend to zero when the number of relays is large. This is a significant difference between traditional methods and the big-data-aided method. For example, if there are 200 relays in the system and the SOP without relay selection is 0.99, then the SOP with relay selection will be $0.99^{200} = 0.134$, which is an intolerable number (too large) for a SOP. However, if the number of relays is 2 000, the value of $0.99^{2000} = 0.18 \times 10^{-8}$ is almost zero. Thus, this shows that the secrecy performance will be improved noticeably by using the big-data-aided method.

5 Hybrid relay network without ECSI

Because it is difficult to obtain the CSI of an eavesdropper in practice, we consider a more practical scene that the ECSI is absent and only its channel distribution information is known in this section. This means that the relay selection is irrelevant to the eavesdropper's capacity. Because the capacity of the destination with the k th relay operating in FD mode is given by Eq. (12), the FD relay selection scheme can be expressed as

$$k_f^* = \arg \max_k C_{d,k}^{\text{FD}}. \quad (40)$$

Similarly, according to the capacity of the destination with the k th relay operating in HD mode by Eq. (20), we obtain the HD relay selection scheme as

$$k_h^* = \arg \max_k C_{d,k}^{\text{HD}}. \quad (41)$$

For the FD mode, the optimal relay can be further expressed as

$$k_f^* = \arg \max_k \min \left\{ \frac{\gamma_{sr_k}}{\gamma_{r_k r_k} + 1}, \gamma_{r_k d} \right\}. \quad (42)$$

Omitting the subscript k , we still use the expression $Y_1 = \min \{ \gamma_{sr} / (\gamma_{rr} + 1), \gamma_{rd} \}$ discussed in the previous section. The variable Y_1^* denotes the maximum Y_1 corresponding to the optimal relay k_f^* . Then the secrecy capacity of the FD relay selection without ECSI is obtained as

$$C_{\text{s-sub}, k_f^*}^{\text{FD}} = \left[\text{lb} \left(\frac{1 + Y_1^*}{1 + \gamma_{se} + \gamma_{re}} \right) \right]^+. \quad (43)$$

Because the channels from the relays to the eavesdropper are i.i.d. and the ECSI is absent, the p.d.f. of $\gamma_{se} + \gamma_{re}$ is irrelevant to relay selection, and can also be expressed by Eq. (31). Thus, the SOP can be obtained as

$$\begin{aligned} P_{\text{out-sub}}^{\text{FD}} &= \Pr \{ C_{\text{s-sub}, k_f^*}^{\text{FD}} < R_s \} \\ &= \Pr \left\{ \text{lb} \left(\frac{1 + Y_1^*}{1 + Z_1} \right) < R_s \right\} \\ &= \int_0^\infty f_{Z_1}(z) F_{Y_1^*} \left(2^{R_s} (1 + z) - 1 \right) dz. \end{aligned} \quad (44)$$

Since the CDF of Y_1 is obtained by Eq. (30), the CDF of Y_1^* can be calculated using

$$\begin{aligned} F_{Y_1^*}(x) &= \left(F_{Y_1}(x) \right)^N \\ &= \left(1 - \frac{e^{-(\lambda_{sr} + \lambda_{rd})x}}{1 + \frac{\lambda_{sr}}{\lambda_{rr}} x} \right)^N, \end{aligned} \quad (45)$$

Combining the p.d.f. of $Z_1 = \gamma_{se} + \gamma_{re}$ in (31), the SOP of the FD relay selection without ECSI can be expressed as

$$\begin{aligned} P_{\text{out-sub}}^{\text{FD}} &= \int_0^\infty \frac{\lambda_{re} \lambda_{se} (e^{-\lambda_{se} z} - e^{-\lambda_{re} z})}{\lambda_{re} - \lambda_{se}} \\ &\quad \cdot \left(1 - \frac{e^{-(\lambda_{sr} + \lambda_{rd})(2^{R_s}(1+z)-1)}}{1 + \frac{\lambda_{sr}}{\lambda_{rr}} (2^{R_s}(1+z)-1)} \right)^N dz. \end{aligned} \quad (46)$$

We now analyze the SOP of the HD relay selection without ECSI. Observing Eq. (20), we find that the optimal relay is given by

$$k_h^* = \arg \max_k \min \{ \gamma_{sr_k}, \gamma_{r_k d} \}. \quad (47)$$

Omitting the subscript k , we still use the expression $Y_2 = \min\{\gamma_{sr}, \gamma_{rd}\}$. We use the notation Y_2^* for the maximum Y_2 corresponding to the optimal HD relay k_h^* . It is clear that the secrecy capacity of the HD relay selection without ECSI is expressed as

$$C_{s_sub, k_h^*}^{HD} = \left[\frac{1}{2} \text{lb} \left(\frac{1 + Y_2^*}{1 + \gamma_{se} + \gamma_{re}} \right) \right]^+, \quad (48)$$

which shows the SOP of HD relay selection without ECSI as

$$\begin{aligned} P_{out_sub}^{HD} &= \Pr\{C_{s_sub, k_h^*}^{HD} < R_s\} \\ &= \Pr\left\{ \frac{1}{2} \text{lb} \left(\frac{1 + Y_2^*}{1 + Z_1} \right) < R_s \right\} \\ &= \int_0^\infty f_{Z_1}(z) F_{Y_2^*} \left(2^{2R_s}(1+z) - 1 \right) dz. \end{aligned} \quad (49)$$

Because the CDF of Y_2 and the p.d.f. of $Z_1 = \gamma_{se} + \gamma_{re}$ are given by Eq. (36) and Eq. (31) respectively, the CDF of Y_2^* can be calculated by using the principle of order statistics random variable, which is given by

$$\begin{aligned} F_{Y_2^*}(x) &= \left(F_{Y_2}(x) \right)^N \\ &= \left(1 - e^{-(\lambda_{sr} + \lambda_{rd})x} \right)^N. \end{aligned} \quad (50)$$

Thus, the concrete result of Eq. (49) is given by

$$\begin{aligned} P_{out_sub}^{HD} &= \sum_{k=0}^N \left[\lambda_{se} \lambda_{re} \binom{N}{k} (-1)^k e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s} - 1)k} \right. \\ &\quad \cdot \left(\lambda_{se} + k(\lambda_{sr} + \lambda_{rd}) 2^{2R_s} \right)^{-1} \\ &\quad \left. \cdot \left(\lambda_{re} + k(\lambda_{sr} + \lambda_{rd}) 2^{2R_s} \right)^{-1} \right], \end{aligned} \quad (51)$$

and the calculation is provided in Appendix C.

It is obvious that an eavesdropper's capacity in HD mode is half that in FD mode from Eq. (16) and Eq. (20), which means that the secrecy capacity in FD mode is possibly less than that in HD mode even though the capacity of the destination in FD mode is larger than that in HD mode. Thus, we introduce a factor ϕ to ensure that the secrecy capacity of the hybrid mode is not less than the arbitrary HD or FD mode, which is expressed as

$$k^* = \arg \max_{k_f^*, k_h^*} \left\{ C_{d, k_f^*}^{FD}, \phi C_{d, k_h^*}^{HD} \right\}. \quad (52)$$

Moreover, it can be further expressed as

$$\begin{cases} \text{HDmode, if } C_{d, k_f^*}^{FD} < \phi C_{d, k_h^*}^{HD}, \\ \text{FDmode, if } C_{d, k_f^*}^{FD} \geq \phi C_{d, k_h^*}^{HD}. \end{cases} \quad (53)$$

The value of ϕ is calculated by the equation $C_{s, k_f^*}^{FD} = C_{s, k_h^*}^{HD}$. However, the capacity of the eavesdropper is unknown. We employ the average capacity instead and the equation is expressed as $\phi C_{d, k_h^*}^{HD} - \bar{C}_{e, k_f^*}^{FD} = C_{d, k_h^*}^{HD} - \bar{C}_{e, k_h^*}^{HD}$. Thus, the value of ϕ is given by

$$\phi = \frac{\bar{C}_{e, k_f^*}^{FD} - \bar{C}_{e, k_h^*}^{HD}}{C_{d, k_h^*}^{HD}} + 1, \quad (54)$$

and the secrecy capacity of hybrid relay mode without ECSI is

$$C_{s_sub, k^*}^{\text{Hybrid}} = \left[\max \left\{ C_{d, k_f^*}^{FD}, \phi C_{d, k_h^*}^{HD} \right\} - C_{e, k^*} \right]^+, \quad (55)$$

where the value of C_{e, k^*} is based on the choice of mode. If it is in HD mode, the expression is given by Eq. (21); if it is in FD mode, the expression is given by Eq. (17). Thus, the SOP of hybrid relay selection without ECSI is given by

$$\begin{aligned} P_{out_sub}^{\text{Hybrid}} &= \Pr\{C_{s_sub, k^*}^{\text{Hybrid}} < R_s\} \\ &= \Pr\left\{ \max \left\{ C_{d, k_f^*}^{FD}, \phi C_{d, k_h^*}^{HD} \right\} < R_s \right\}. \end{aligned} \quad (56)$$

Now, we analyze the SOP from a big-data perspective. It is clear that the value of the function in Eq. (45) tends to zero when we have $N \rightarrow \infty$. Since $\int_0^\infty f_{Z_1}(z) dz = 1$ is true, the SOP of the FD relay selection in Eq. (46) also tends to zero when $N \rightarrow \infty$ is satisfied. However, the rate of its decline will be obviously slower than that with ECSI (exponential), which is shown in section 6. Similarly, the SOP of the HD and hybrid relay selection scheme will also tend to zero with $N \rightarrow \infty$. This indicates that the secrecy even without ECSI will also be improved significantly with the aid of big-data when there is a large number of relays in the system.

6 Simulation results

In this section, we present simulation results to evaluate the SOP of HD, FD, and hybrid relay selection

Figure 2 SOP vs. channel SNR. (a) SOP with ECSI vs. channel SNR ($N=200$), $\sigma_{sr}^2 = \sigma_{rd}^2$, $\sigma_{se}^2 = 20$ dB, $\sigma_{re}^2 = 25$ dB, $\sigma_{rr}^2 = 5$ dB; (b) SOP without ECSI vs. channel SNR ($N = 200$), $\sigma_{sr}^2 = \sigma_{rd}^2$, $\sigma_{se}^2 = 15$ dB, $\sigma_{re}^2 = 20$ dB, $\sigma_{rr}^2 = 5$ dB.

Figure 3 SOP vs. self-interference. (a) SOP ECSI vs. Self-interference ($N=200$), $\sigma_{sr}^2 = \sigma_{rd}^2 = 30$ dB, $\sigma_{se}^2 = 20$ dB, $\sigma_{re}^2 = 25$ dB; (b) SOP vs. Self-interference ($N=200$), $\sigma_{sr}^2 = \sigma_{rd}^2 = 30$ dB, $\sigma_{se}^2 = 10$ dB, $\sigma_{re}^2 = 15$ dB.

schemes. There are several common assumptions included in the simulation: the length of successive codewords is $L = 10$, the destination and the relay have the same channel variance, i.e., $\sigma_{sr}^2 = \sigma_{rd}^2$, and the transmit power at the source and relay is $p_s = p_r = 1$.

The SOPs with and without ECSI versus main channel SNR are depicted in Fig. 2. It is clear that the simulation results match theoretical results well. When self-interference is obviously weaker than the main channel SNR, the SOP of the hybrid relay selection scheme with known ECSI is better than that of the HD and FD relay selection schemes, but it is

the same as the FD relay selection scheme when the ECSI is unknown because the secrecy capacity of the FD mode is always larger than that of the HD mode. In this case, the performance of the HD mode is significantly worse than the others.

Fig. 3 shows the SOPs of HD, FD, and hybrid relay selection schemes versus the self-interference of FD relay. It is obvious that a break-even value exists. When the self-interference is less than this value, the performance of the FD mode is better than that of the HD mode. Otherwise, the HD mode is better than the FD mode. The hybrid relay selection scheme overcomes this problem and exhibits superi-

Figure 4 SOP vs. the number of relay. (a) SOP vs. Number of relay, $\sigma_{sr}^2 = \sigma_{rd}^2 = 30$ dB, $\sigma_{se}^2 = 20$ dB, $\sigma_{re}^2 = 25$ dB, $\sigma_{rr}^2 = 5$ dB; (b) SOP without ECSI vs. Number of relay, $\sigma_{sr}^2 = \sigma_{rd}^2 = 30$ dB, $\sigma_{se}^2 = 15$ dB, $\sigma_{re}^2 = 20$ dB, $\sigma_{rr}^2 = 5$ dB.

Figure 5 SOP vs. the target secrecy rate. (a) SOP with ECSI vs. R_s , $\sigma_{sr}^2 = \sigma_{rd}^2 = 30$ dB, $\sigma_{se}^2 = 15$ dB, $\sigma_{re}^2 = 20$ dB, $\sigma_{rr}^2 = 10$ dB, $N = 100$; (b) SOP without ECSI vs. R_s , $\sigma_{sr}^2 = \sigma_{rd}^2 = 30$ dB, $\sigma_{se}^2 = 10$ dB, $\sigma_{re}^2 = 15$ dB, $\sigma_{rr}^2 = 10$ dB, $N = 100$.

ority. At the same time, Fig. 3(b) shows that the performance of the hybrid relay selection scheme is almost the same as that of the FD mode, meaning that the performance of the hybrid relay selection scheme is mainly decided by the FD mode (i.e., the self-interference) when the eavesdropper's CSI is unknown.

The change of SOPs with the number of relays N is shown in Fig. 4. The decrease of SOP with ECSI is linear, exhibited much faster than that without ECSI. The performance of the FD mode and hybrid mode is much better than that of the HD mode when the number of relays is large. It is clear that the SOP can be controlled as a really small value when

the number of relays is large enough even though the eavesdropper can intercept the private message at two phases. For example, the figure shows that the SOP of the FD scheme is 10^{-20} when $N=2000$ in the case of having ECSI, and the SOP of the FD scheme is less than 10^{-5} when $N > 6000$ in the case of no ECSI, which is obviously better than the case with $N = 200$. This result shows that the secrecy can be improved significantly in big-data era when the number of relays is large.

Fig. 5 shows the SOPs versus the target secrecy rate. As indicated from Fig. 5(a), when the target secrecy rate is small, the SOP of the HD relay selection scheme is better than the FD relay selection scheme.

Up to the point that the target secrecy rate is larger than a break-even value (about $1.65 \text{ bit} \cdot \text{s}^{-1}/\text{Hz}$), the FD mode is better than the HD mode. The hybrid technique retains superiority regardless of the target secrecy rate. However, the situation changes when the number of relays becomes large under the scenario that the ECSI is unknown. The HD mode is far worse than the FD mode when $N = 100$, although it is better than the FD mode when $N = 10$ at the point $R_s = 1.67 \text{ bit} \cdot \text{s}^{-1}/\text{Hz}$. This means that a large number of relays may change the result of mode selection in some cases.

7 Conclusion

In this paper, the SOP of a big-data-aided hybrid relay selection scheme has been analyzed. With ECSI, the performance of the hybrid relay selection scheme was always better than that of the HD and FD relay selection schemes. However, it is almost the same as the FD mode without eavesdropper's CSI when self-interference is low, meaning that the performance of the hybrid relay selection scheme is mainly decided by the self-interference of the FD technique. From the simulation results, it is clear that the secrecy constraint can be guaranteed in a rigorous scope when the number of relays is large enough even though the eavesdropper has the advantage of two-phases' eavesdropping. Obviously, the extra overhead caused by the increased number of relays is resolvable in big-data era. Finally, the large number of relays may change the result of mode selection in some cases, meaning a possible difference between traditional designs and big-data-aided design.

Appendix

A) Because $Y_1 = \min\{\frac{\gamma_{sr}}{\gamma_{rr} + 1}, \gamma_{rd}\}$, $Z_1 = \gamma_{se} + \gamma_{re}$, and they are independent, the distribution of $X_2 = \text{lb}\left((1 + Y_1)/(1 + Z_1)\right)$ is given by

$$\begin{aligned} F_{X_2}(x) &= \Pr(X_2 < x) = \Pr\left(\text{lb}\left(\frac{1 + Y_1}{1 + Z_1}\right) < x\right) \\ &= \int_0^\infty \int_0^{2^{2^x(1+z)}-1} f_{Y_1}(y) f_{Z_1}(z) dy dz \end{aligned}$$

$$\begin{aligned} &= \int_0^\infty f_{Z_1}(z) F_{Y_1}\left(2^x(1+z) - 1\right) dz \\ &= 1 - ma \int_0^\infty \frac{e^{-(\lambda_{se}+b)z} - e^{-(\lambda_{re}+b)z}}{c + dz} dz \\ &= 1 - \frac{ma \cdot e^{\frac{(\lambda_{se}+b)c}{d}}}{d} \int_{\frac{(\lambda_{se}+b)c}{d}}^\infty \frac{e^{-t}}{t} dt \\ &\quad + \frac{ma \cdot e^{\frac{(\lambda_{re}+b)c}{d}}}{d} \int_{\frac{(\lambda_{re}+b)c}{d}}^\infty \frac{e^{-t}}{t} dt \\ &= 1 - \frac{ma \cdot e^{\frac{(\lambda_{se}+b)c}{d}}}{d} E_1\left(\frac{(\lambda_{se}+b)c}{d}\right) \\ &\quad + \frac{ma \cdot e^{\frac{(\lambda_{re}+b)c}{d}}}{d} E_1\left(\frac{(\lambda_{re}+b)c}{d}\right). \quad (57) \end{aligned}$$

where $E_1(x) = \int_x^\infty e^{-t} t^{-1} dt$ denotes the exponential integral function, $m = \lambda_{re}\lambda_{se}/(\lambda_{re} - \lambda_{se})$, $a = \exp\{-(\lambda_{sr} + \lambda_{rd})2^x + \lambda_{sr} + \lambda_{rd}\}$, $b = (\lambda_{sr} + \lambda_{rd})2^x$, $c = 1 + (\lambda_{sr}/\lambda_{rr})2^x - \lambda_{sr}/\lambda_{rr}$, and $d = 2^x \cdot \lambda_{sr}/\lambda_{rr}$.

B) Because $Y_2 = \min\{\gamma_{sr}, \gamma_{rd}\}$, $Z_1 = \gamma_{se} + \gamma_{re}$ are independent, the distribution of $X_3 = (1/2) \cdot \text{lb}\left((1 + Y_2)/(1 + Z_1)\right)$ is expressed as

$$\begin{aligned} &F_{X_3}(x) \\ &= \Pr(X_3 < x) = \Pr\left(\frac{1}{2} \text{lb}\left(\frac{1 + Y_2}{1 + Z_1}\right) < x\right) \\ &= \int_0^\infty \int_0^{2^{2^x(1+z)}-1} f_{Y_2}(y) f_{Z_1}(z) dy dz \\ &= \int_0^\infty f_{Z_1}(z) F_{Y_2}\left(2^{2^x(1+z)} - 1\right) dz \\ &= 1 - mq \int_0^\infty \left(e^{-(\lambda_{se}+p)z} - e^{-(\lambda_{re}+p)z}\right) dz \\ &= 1 - mq \cdot \left(\frac{1}{\lambda_{se} + p} - \frac{1}{\lambda_{re} + p}\right) \\ &= 1 - \frac{\lambda_{re}\lambda_{se} \cdot q}{(\lambda_{se} + p)(\lambda_{re} + p)}, \quad (58) \end{aligned}$$

where $m = \lambda_{re}\lambda_{se}/(\lambda_{re} - \lambda_{se})$, $p = (\lambda_{sr} + \lambda_{rd}) \cdot 2^{2^x}$, and $q = \exp\{-(\lambda_{sr} + \lambda_{rd}) \cdot (2^{2^x} - 1)\}$.

C) In this appendix, we provide the calculation of Eq. (51):

$$\begin{aligned} &\int_0^\infty f_{Z_1}(z) F_{Y_2^*}\left(2^{2R_s}(1+z) - 1\right) dz \\ &= \int_0^\infty m e^{-\lambda_{se}z} \cdot \left(1 - e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s}(1+z)-1)}\right)^N dz \end{aligned}$$

$$-\int_0^\infty m e^{-\lambda_{re} z} \cdot \left(1 - e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s}(1+z)-1)}\right)^N dz.$$

where $m = \lambda_{re} \lambda_{se} / (\lambda_{re} - \lambda_{se})$. We start by calculating the first item of the Eq. (59),

$$\begin{aligned} & \int_0^\infty m e^{-\lambda_{se} z} \cdot \left(1 - e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s}(1+z)-1)}\right)^N dz \\ &= \int_0^\infty m e^{-\lambda_{se} z} \sum_{k=0}^N \binom{N}{k} (-1)^k \\ & \quad \cdot e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s}(1+z)-1) \cdot k} dz \\ &= \sum_{k=0}^N m \binom{N}{k} (-1)^k e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s}-1)k} \\ & \quad \cdot \int_0^\infty e^{-\mu_k z} dz \\ &= \sum_{k=0}^N \frac{m \binom{N}{k} (-1)^k e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s}-1)k}}{\lambda_{se} + k(\lambda_{sr} + \lambda_{rd})2^{2R_s}}, \end{aligned} \quad (59)$$

where $\mu_k = \lambda_{se} + k(\lambda_{sr} + \lambda_{rd})2^{2R_s}$. Similarly, the second item of Eq. (59) is given by

$$\begin{aligned} & \int_0^\infty m e^{-\lambda_{re} z} \cdot \left(1 - e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s}(1+z)-1)}\right)^N dz \\ &= \sum_{k=0}^N \frac{m \binom{N}{k} (-1)^k e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s}-1)k}}{\lambda_{re} + k(\lambda_{sr} + \lambda_{rd})2^{2R_s}}. \end{aligned} \quad (60)$$

Thus, we obtain the result

$$\begin{aligned} & \int_0^\infty f_{Z_1}(z) F_{Y_2^*} \left(2^{2R_s}(1+z)-1\right) dz \\ &= \sum_{k=0}^N \left[\lambda_{se} \lambda_{re} \binom{N}{k} (-1)^k e^{-(\lambda_{sr} + \lambda_{rd})(2^{2R_s}-1)k} \right. \\ & \quad \cdot \left(\lambda_{se} + k(\lambda_{sr} + \lambda_{rd})2^{2R_s} \right)^{-1} \\ & \quad \cdot \left. \left(\lambda_{re} + k(\lambda_{sr} + \lambda_{rd})2^{2R_s} \right)^{-1} \right]. \end{aligned} \quad (61)$$

For the above expressions, we have used the binomial theorem $(a+b)^N = \sum_{k=0}^N \binom{N}{k} a^{N-k} b^k$.

References

- [1] V. N. Q. Bao, T. N. Linh, M. Debbah. Relay selection schemes for dual-hop networks under security constraints with multiple eavesdroppers [J]. IEEE transactions wireless communications, 2013, 12(12): 6076-6085.
- [2] Y. Liu, L. Wang, T. T. Duy, et al. Relay selection for security enhancement in cognitive relay networks [J]. IEEE wireless communications letters, 2015, 4(1): 46-49.
- [3] G. Chen, Z. Tian, Y. Gong, et al. Max-ratio relay selection in secure buffer-aided cooperative wireless networks [J]. IEEE transactions on information forensics and security, 2014, 9(4): 719-729.
- [4] I. Krikidis, J. Thompson, S. Mclaughlin. Relay selection for secure cooperative networks with jamming [J]. IEEE transaction wireless communications, 2009, 8(10): 5003-5011.
- [5] I. Krikidis. Opportunistic relay selection for cooperative networks with secrecy constraints [J]. IET communications, 2010, 4(15): 1787-1791.
- [6] Y. L. Zou, X. B. Wang, W. M. Shen. Optimal relay selection for physical-layer security in cooperative wireless networks [J]. IEEE journal on selected areas in communications, 2013, 31(31): 2099-2111.
- [7] I. Krikidis, H. A. Suraweera, C. Yuen. Amplify-and-Forward with full-duplex relay selection [C]//IEEE International Conference on Communications, 2012: 1-6.
- [8] B. Zhong, Z. Zhang, X. Chai, et al. Performance analysis for opportunistic full-duplex relay selection in underlay cognitive networks [J]. IEEE transactions on vehicular technology, 2015, 64(10): 4905-4910.
- [9] L. Lai, H. E. Gamal. The relay-eavesdropper channel: cooperation for secrecy [J]. IEEE transactions on information theory, 2008, 54(9): 4005-4019.
- [10] L. Dong, Z. Han, A. Petropulu, et al. Improving wireless physical layer security via cooperating relays [J]. IEEE transactions on signal processing, 2010, 58(3): 1875-1888.
- [11] J. Li, A. P. Petropulu, S. Weber. On cooperative relaying schemes for wireless physical layer security [J]. IEEE transactions on signal processing, 2011, 59(10): 4985-4997.
- [12] D. S. Michalopoulos, J. Schlenker, J. Cheng, et al. Error rate analysis of full-duplex relaying [C]//International Waveform Diversity and Design Conference, 2010: 165-168.
- [13] Z. Ding, K. K. Leung, D. L. Goeckel, et al. On the application of cooperative transmission to secrecy communications [J]. IEEE journal on selected areas in communications, 2012, 30(2): 359-368.
- [14] S. Bi, R. Zhang, Z. Ding, et al. Wireless communications in the era of big data [J]. IEEE communications magazine, 2011, 53(10): 190-199.
- [15] K. Yang, Q. Yu, S. Leng, et al. Data and energy integrated communication networks for wireless big data [J]. IEEE access, 2016, 4: 713-723.
- [16] A. A. Crdenas, P. K. Manadhata, S. P. Rajan. Big data analytics for security [J]. IEEE security & privacy, 2013, 11(6): 74-76.
- [17] H. Alves, G. Brante, R. D. Souza, et al. On the performance of secure full-duplex relaying under composite fading channels [J]. IEEE signal processing letters, 2015, 22(7): 867-870.
- [18] H. Alves, G. Brante, R. Souza, et al, Latva-aho M. On

the performance of full-duplex relaying under physical security constraints [C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014: 1-5.

- [19] J. H. Lee. Full-duplex relay for enhancing physical layer security in multi-Hop relaying systems [J]. IEEE communications letters, 2015, 22(4): 525-528.
- [20] G. Chen, Y. Gong, P. Xiao, et al. Physical layer network security in the full-duplex relay system [J]. IEEE transactions on information forensics and security, 2015, 10(3): 574-583.
- [21] S. Parsaefard, T. Le-Ngoc. Improving wireless secrecy rate via full-duplex relay-assisted protocols[J]. IEEE transactions on information forensics and security, 2015, 10(10): 2095-2107.
- [22] T. Riihonen, S. Werner, R. Wichman. Hybrid full-duplex/halfduplex relaying with transmit power adaptation [J]. IEEE transactions on wireless communications, 2011, 10(9): 3074-3085.
- [23] J. Lee, T. Q. S. Quek. Hybrid full-/half-duplex system analysis in heterogeneous wireless networks [J]. IEEE transactions on wireless communications, 2015, 14(5): 2883-2895.
- [24] P. Lioliou, M. Viberg, M. Coldrey, et al. Self-interference suppression in full-duplex MIMO relays [C]//Asilomar Conference on Signals, Systems and Computers Pacific Grove, 2007: 658-662.
- [25] T. Riihonen, S. Werner, R. Wichman, Mitigation of loopback self-interference in full-duplex MIMO relays [J]. IEEE transactions on signal processing, 2011, 59(12): 5983-5993.
- [26] T. Riihonen, S. Werner, R. Wichman. Optimized gain control for single-frequency relaying with loop interference [J]. IEEE transactions wireless communications, 2009, 8(6): 2801-2806.
- [27] W. C. Yueh, S. S. Cheng. Explicit eigenvalues and inverses of several toeplitz matrices [J]. Anziam Journal, 2006, 48: 73-97.
- [28] A. Papoulis. Probability, random variables, and stochastic processes [M], 4th ed. New York: McGraw-Hill, 2002.

Pinyi Ren [corresponding author] received his B.S., M.S. and Ph.D. degrees from Xi'an Jiaotong University. He is now a professor of information and communications engineering, Xi'an Jiaotong University. His research interests include cognitive radio networks, MIMO systems, game theory in wireless communications, wireless relay, routing, and signal detection. (Email: pyren@mail.xjtu.edu.cn)

Qinghe Du received his Ph.D. from Texas A&M University, USA. He is now an associate professor of Information and Communications Engineering Department, Xi'an Jiaotong University. His research interests include mobile wireless communications and networking with an emphasis on mobile multicast, statistical QoS (Quality-of-Service) provisioning, and cognitive radio networks. (Email: duqinghe@mail.xjtu.edu.cn)

Li Sun received his Ph.D. from Xi'an Jiaotong University, China. He is now an associate professor of information and communications engineering, Xi'an Jiaotong University. His research interests include cooperative relaying networks, wireless physical-layer security, and device-to-device communications. (Email: lisun@mail.xjtu.edu.cn)

Yichen Wang received his Ph.D. from Xi'an Jiaotong University, China. He is now an associate professor of information and communications engineering, Xi'an Jiaotong University. His research interests include mobile wireless communications and networks with an emphasis on cognitive radio techniques, ad hoc networks, MAC protocol design, statistical QoS provisioning, and resource allocation over wireless networks. (Email: wangyichen0819@mail.xjtu.edu.cn)

About the authors

Hongliang He received the B.E. degree in information and communication engineering from Harbin Engineering University. He is now a Ph.D. candidate in Xi'an Jiaotong University. His research interests include wireless physical-layer security, Internet of Things and cooperative communications. (Email: hehongliang@stu.xjtu.edu.cn)

Hongliang He received the B.E. degree in information and communication engineering from Harbin Engineering University. He is now a Ph.D. candidate in Xi'an Jiaotong University. His research interests include wireless physical-layer security, Internet of Things and cooperative communications. (Email: hehongliang@stu.xjtu.edu.cn)