IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 6, 2020, accepted May 30, 2020, date of publication June 2, 2020, date of current version June 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2999450

Outage Analysis and Learning-Based Relay Selection for Opportunistic Lossy Forwarding Relaying Systems

HAIYANG ZHANG[1] AND ZHIWEI ZHANG[2]
[1]School of Environment and Geomatics Engineering, Suzhou University, Suzhou 234000, China
[2]School of Informatics and Engineering, Suzhou University, Suzhou 234000, China

Corresponding author: Haiyang Zhang (hy_zhang_sz@163.com)

This work was supported in part by the Key Project of Natural Science Foundation of the Department of Education of Anhui Province under Grant KJ2018A0449 and Grant KJ2014ZD31, in part by the Key Project of Education Quality of Anhui Province under Grant 2017jyxm0510, and in part by the Anhui Provincial Science Foundation under Grant 1908085QF283.

ABSTRACT In this paper, we study an opportunistic decode-and-forward (DF) system allowing intra-link errors, referred as opportunistic lossy-forwarding (OLF) system. Multiple relays are available but only one ‘‘best relay’’ is selected to forward the decoded information to the destination. The forwarded information may be erroneous. The relay selection strategy is built using source coding with a helper theorem. We evaluate the outage probability of OLF relaying system, and derive an approximated theoretical outage probability bound. For practical deployment, a statistical learning algorithm for selection relay is designed. The practical performance loss is negligible compared to the optimal selection case. The OLF system outperforms the existing opportunistic DF system in terms of outage performance. Both theoretical and numerical results show that full diversity can be achieved. Finally, we investigate the position of deploying relays by minimizing the outage probability from simulation results.

INDEX TERMS Cooperative communications, relaying protocol, lossy forwarding, opportunistic relay, diversity order, outage performance.

I. INTRODUCTION

Cooperative wireless communication, especially relay assisted communication, has been gaining popularity in wireless sensor networks to enhance the network performance in terms of coverage area, energy consumption and throughput [1], [2]. With multiple geographically dispersed nodes, cooperative wireless systems may take more advantages in multipath diversity than multiple input multiple output (MIMO) systems to get rid of oversize and power-hungry antenna array on the power and size limited wireless nodes. However, tackling multiple relays at the same time in practice is quite trivial particularly when the number of relays is relatively large [3]. Various associated topics have been investigated especially distributed space-time coding techniques [4]–[6].

By activating the whole set of relays and employing maximal-ratio combining (MRC) at the destination seems to be a feasible solution to achieve the full diversity, however, it suffers from high decoding complexity and power consumption. To alleviate this problem, both one-way and two-way opportunistic relaying have been widely studied [3], [7]–[11]. Instead of activating the whole set of relays, it provides relatively simple methods to achieve cooperative diversity by selecting a ‘‘best’’ relay. The selection criteria is based on the quality of the source-relay link or source-relay-destination link. One-way opportunistic relaying has been first proposed in [3], where a distributed relay selection strategy has been shown that full diversity can still be achieved. A closed-form expression for the outage probability of decode-and-forward (DF) with selective combining has been derived in [12], which shows that selecting the best relay has only caused 1 dB performance loss compared activating the whole set of relays. Two-way relaying is also an important topic for opportunistic selection. It has been proven that two-way relaying may also accomplish full diversity order with network coding [9].

However, since errors/noises of the source-relay link can propagate in the destination that may affect the achievable

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

diversity [9], [13], source-relay link (referred as intra-link) errors are usually not allowed in the state-of-the-art opportunistic relaying systems, especially for DF relaying systems which are widely adopted in cooperative communication systems. In the future cooperative wireless network scenarios, intra-link error can be inevitable. For example, in 5G V2X networks, channels of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) can be very unstable due to fast moving objects. It has been shown that lossy-forwarding relaying system [14] is able to achieve better performance than the existing DF relaying system, by allowing the intra-link errors and performing joint decoding at the destination.

Recently, machine learning has attracted interests in many disciplines as well as the wireless communication area [15]. In [16], [17], the authors have proposed a novel lookup table (LUT) decoding method for low-density parity check codes (LDPC) based on statistical learning theory, which is a typical application of supervised learning. The result has shown that the low complexity LUT method outperformed two conventional quantized decoding algorithms. In [18], the clustering theory, which is a typical application for unsupervised learning, has been modeled by universal communication in the perspective of information theory.

In this paper, motivated by an emerging DF system allowing intra-link errors, referred as lossy-forwarding (LF), and the statistical learning methods for wireless communications, we propose an opportunistic lossy-forwarding (OLF) relaying system. The main contributions of this paper are summarized as follows.

- We propose an opportunistic relay selection strategy for the multi-relay lossy-forwarding system, and derived an approximated theoretical outage probability bound when an arbitrary relay set is activated or selected.
- We also propose a practical relay selection method based on a supervised learning model to select the "best" relay in the lossy-forwarding relaying framework. Different from the DF relaying, we consider that the qualities of both source-relay and relay-destination links have to be equally taken account when selecting the relay. In practice, since the scenarios may vary, depending on the signal parameters as well as propagation scenarios, we adopt a statistical learning based method for any specified implementation of our system. It should be emphasized that the proposed technique can also be utilized for some other opportunistic node selection methods.
- We validate the performance of the proposed OLF relaying system and the relay selection algorithm through conducted simulations, which show the consistency between the theoretical and simulation results. Numerical results also indicate that our system outperformed the existing opportunistic DF system both in theory and practice with full diversity gain.

The rest of this paper is organized as follows. Section II demonstrates the OLF relaying system model with location settings. The exact outage probability and its approximated theoretical bound for our OLF system are proposed in Section III. We present the theoretical relay selection strategy and our designed machine learning-based relay selection algorithm in Section IV. Section V shows the numerical results obtained by conducted simulations. We give some discussions on the numerical results in Section VI. We finally conclude our work in Section VII.

II. SYSTEM MODEL

Figure 1 illustrates the OLF relaying system model which is investigated by us. The system consists of a source node S, a destination node D and a set of relays. We use $\mathbb{R}$ to denote the whole set of relay nodes, and R_i with $i \in \{1, \ldots, m\}$ to represent each relay node in the set. All the communication links are assumed to be independent block Rayleigh fading channels, denoted by h_{SD}, h_{SR_i} and h_{R_iD} as the channel gains respectively. The channel gain follows complex Gaussian distribution with mean 0 and variance 1. Hence, for each link, the receive signal can be expressed as $y = h \cdot s + n$, where s is the signal to be transmitted by each node, and n is the additive white Gaussian noise with variance $\sigma^2 = N_0/2$ and mean 0. We consider three typical locations for examining the outage performance, which is illustrated in Fig. 2. For simplicity, we consider the relay node is located near the source, and in the centre between the source and destination nodes. Besides, we also evaluate the system performance by moving the relay node from the source to the destination.

The whole transmission has $2 + \epsilon$ timeslots. In the first timeslot, the source node S transmits its message to other

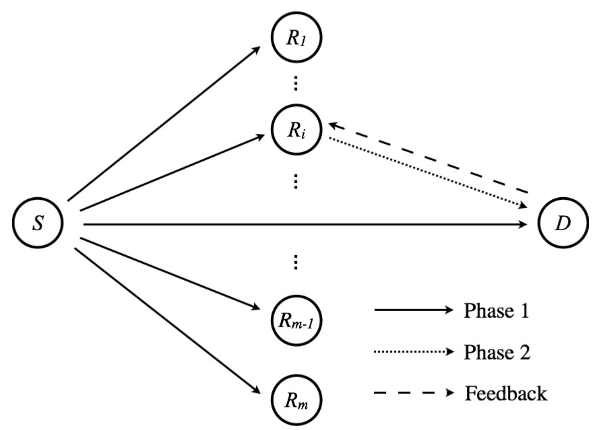

FIGURE 1. An OLF system model with m relay nodes.

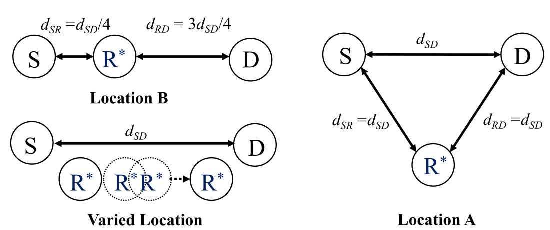

FIGURE 2. Location settings in our OLF relaying system, where R^* stands for the selected relay.

nodes via a broadcast channel. In the second timeslot, only one relay node R_i is selected and activated to forward its message to assist the *SD* transmission, based on a designed selection criterion. Due to the fact that the intra-link errors are allowed in LF systems, the selection of the relay R_i is performed without regard to successful decoding. The destination node D performs joint decoding with the signals transmitted from both S and the selected R_i. The selection is determined by the destination node D through an ϵ feedback timeslot. We assume that the channel state information (CSI) of each link is known by the destination. Using ϵ to indicate the feedback slot is because the required transmission resources of feedback information is negligible compared to that of the information.

A lossy-forwarding system is thus formed by the source S, the destination D and the selected relay R_i, as in [19], which is shown in Fig. 3. The source message sequence $\mathbf{u}_S$ is encoded by a serial concatenated convolutional codes (SCCC) encoder, typically including an outer convolutional encoder, an interleaver, an inner accumulator and a modulator. Without loss of generality, we assume that the spectrum efficiency (including the total coding rate and the modulation order) is R_c. Once the signal is detected at the relay, it decodes the received signal simply and generate an estimation $\hat{\mathbf{u}}_S$. The estimation is then interleaved by a random interleaver π, followed by a SCCC encoder with different parameters. After that, the re-encoded sequence goes through a modulator and its output signal is transmitted to the destination D. At the destination, the signals from the source and the relay are fed into a joint decoder which takes the benefit of the correlation between the signals. In particular, function f_c [20] associates different weights to the extrinsic log-likelihood $\mathbf{L}_e^u$ according to the correlation level, i.e., if the correlation is high, a large weight is assigned to the extrinsic log-likelihood, and vice versa. This encoding and decoding structure is adopted in the simulations and generating the training data for relay selection.

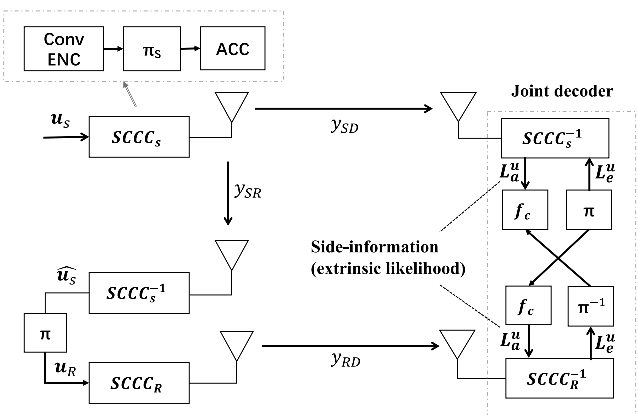

FIGURE 3. The system model of a typical three-node lossy-forwarding relaying system. Conv ENC represents convolutional code encoder, and ACC denotes accumulator.

III. THEORETICAL ANALYSES FOR OPPORTUNISTIC LOSSY-FORWARDING

In this section, we will provide the definition of the "best relay" and analyze the outage probability for the OLF system. We start from the formal definition of the outage event and turn to derive an approximated lower bound on the outage probability based on the probability theory.

A. OUTAGE PROBABILITY DEFINITION

An opportunistic relaying system is to find a relay node and activate it to help the transmission, with the aim of minimizing the outage probability. In order to mathematically define the outage probability, we first give a formal definition on the best relay R^* for each transmission, which is summarized as below

Definition 1: If a transmission outage happens with the selected R^*, then selecting any relay R_j from the set $\{R_1, \ldots, R_m\} \backslash R^*$ must cause a transmission outage.

It should be emphasized here that even with the selected best relay, the transmission may fail, i.e., the best relay does not guarantee the successful decoding. Besides, there may exist a group of *best* relays R^*. In fact, higher the signal-to-noise ratio (SNR), the more number of relays will be selected. For simplicity, we always adopt only one best relay R^* by randomly taking one from a set of best relays. Such random selection does not affect the outage performance because that the selection is based on the same metric. For example, in the sufficiently high SNR regime, any relay from the candidates is able to achieve successful decoding, i.e., the outage will not happen.

Intuitively, we can select the relay node one-by-one and forward its information to the destination node to assist the *SD* communication. If none of the selected relay node is able to help the decoding process, then the outage should happen. Hence, according to Definition 1, the outage event can be considered that the transmission fails if all relays have been selected respectively, with the probability being defined as

$$\Pr(E_{R^*}) = \Pr(E_{R_1} \cap E_{R_2} \ldots E_{R_3} \cap E_{R_m}), \tag{1}$$

where E_{R^*} is an event that the joint decoding process at the destination fails (corresponding to an outage) with the selected best relay node R^*, E_{R_i} is an event that relay R_i is selected and the joint decoding fails.

However, it is extremely difficult to derive the closed-form of the outage probability. The reason is two-fold: (1) By the state-of-the-art, outage probability analysis for E_{R_i} (single relay's case) has only been proposed in [21], where they have derived the outage probability. The outage probability is given by a triple integrals, which correspond to each link. In the m-relay's cases, at least $2m + 1$ multiple definite integrals are necessary to obtain the outage probability. The computation is not tackleable even when m is not large. (2) Even more difficult thing is to determine the lower and upper limits from the correlation model of these relays, since it is not a simple task to build the correlation model and derive the distribution function of the channel for the *SR* link.

Instead, we adopt Monte Carlo simulation to calculate the exact outage probability in (1), and derive an approximated bound of (1).

B. OUTAGE PROBABILITY APPROXIMATION

We now turn to examine an approximation of the outage probability using Bayes's theorem. The approximated bound gives the lower bound on the exact outage probability and can be adopted to assist the practical deployment of the OLF relaying system.

Since all the events E_{R_i} are dependent, we cannot calculate (1) by simply multiplying each probability $\Pr\{E_{R_i}\}$. Notice that the correlations among E_{R_i} are totally depended on the direct link *SD*. We can estimate the correlated event by introducing the outage event E_{SD} of *SD* link as a conditional event. By the Rayleigh fading assumption, the outage probability of *SD* is expressed as

$$\Pr(E_{SD}) = \int_0^{\gamma_{th1}} p(\gamma)d\gamma, \tag{2}$$

$$= \int_0^{\gamma_{th1}} \frac{1}{\Gamma_{SD}} \exp(-\frac{\gamma}{\Gamma_{SD}})d\gamma, \tag{3}$$

$$= 1 - \exp(-\frac{\gamma_{th1}}{\Gamma_{SD}}), \tag{4}$$

where γ_{th1} represents the threshold SNR of *SD* link at the destination. The average and instantaneous SNR of *SD* link are Γ_{SD} and γ respectively. The probability density function (pdf) is denoted by $p(\gamma)$ which is assumed to follow Rayleigh distribution.

Due to the difficulty of directly calculating the joint probability $\Pr(E_{R_1} \cap E_{R_2} \ldots E_{R_3} \cap E_{R_m})$ in (1), the Bayes' theorem is applied to decompose the expression as,

$$\Pr(E_{R^*}) \approx \frac{\Pr[\bigcap_{R_i\in\mathbb{R}}(E_{R_i})|E_{SD}] \times \Pr(E_{SD})}{\Pr[E_{SD}|\bigcap_{R_i\in\mathbb{R}}(E_{R_i})]} \tag{5}$$

$$= \Pr[\bigcap_{R_i\in\mathbb{R}}(E_{R_i})|E_{SD}] \times \Pr(E_{SD})$$

$$= \prod_{R_i\in\mathbb{R}} \Pr(E_{R_i}|E_{SD}) \times \Pr(E_{SD}) \tag{6}$$

$$= \frac{\prod_{R_i\in\mathbb{R}} \Pr(E_{R_i}, E_{SD})}{\Pr^{m-1}(E_{SD})}$$

$$= \frac{\prod_{R_i\in\mathbb{R}} \Pr(E_{R_i})}{\Pr^{m-1}(E_{SD})}, \tag{7}$$

where E_{SD} is introduced as a necessary condition of E_{R_i}, and explanations of some key steps are as follows.

(5): the Bayes' theorem is applied;

(6): given the fact that the *SD* link transmission fails, the events that signals coming from all *RD* links fail to contribute to a successful joint decoding are independent. Therefore, the joint probability of $\Pr[\bigcap_{R_i\in\mathbb{R}}(E_{R_i})|E_{SD}]$ can be expressed by a multiplication of individual probabilities.

(7): since E_{SD} is a necessary condition of E_{R_i}, E_{SD} is a subevent of E_{R_i}, and their joint probability equals to $\Pr(E_{R_i})$.

Here, $\Pr(E_{R_i})$ is the lossy-forwarding outage probability with the relay R_i and provided in [21]. For completeness of this paper, we give a short explanation of $\Pr(E_{R_i})$ in Appendix A and its expression in Appendix B. Therefore, the theoretical lower bound on the outage probability of the OLF relaying system is given by

$$\Pr(E_{R^*}) = \frac{\prod_{R_i\in S} \Pr(E_{R_i})}{\Pr^{m-1}(E_{SD})}. \tag{8}$$

Remark: The outage probability derived using (8) is indeed an approximation. To derive the exact bound, we need first derive the channel distribution of SR^* and R^*D links where R^* is the optimal relay node selected by a selection algorithm, i.e., the probability density function of channel SNR's γ_{SR^*} and γ_{R^*D}. Then, by substituting the probability density functions into $\Pr(E_{R_i})$ of lossy-forwarding system [21] to obtain the exact bound. However, deriving the probability density functions is extremely difficult. The reason is due to the optimal selection criterion in the LF system (detailed in Section IV) has a complicated form. Instead, to avoid deriving the joint *pdf* of the channels, we introduce the event E_{SD} to separate the events $E_{R_i}|E_{SD}$. Therefore, we use the $\approx$ sign in (5) to indicate it is an approximation bound (lower bound) on the outage probability.

Remark: There exists a gap between the exact outage bound (1) and the approximated bound (8). However, the approximated bound gives us the opportunities to examine the behavior of the OLF system simply, especially, when the number of relays m gets large. Also, we can use this approximated bound to build optimization algorithms to improve the system performance, for instance, the relay deployment [21], [22], power allocation [23]–[27].

Remark: An interesting thing is to evaluate the diversity order which the system can achieve. We found that the OLF relaying system with m relays can achieve $m + 1$ order diversity. A proposition and its proof are given blow.

Proposition 1: The opportunistic LF system with m relays can achieve $m + 1$ order diversity gain.

Proof (Proof of Proposition 1): Based on the estimation method in [26], [28], in high SNR region, $\Pr(E_{R_i})$ is approximated as

$$\Pr(E_{R_i}) \approx \frac{c_1}{\Gamma_{SD}^2} + \frac{c_2}{\Gamma_{SR_i}^2} + \frac{c_3}{\Gamma_{SR_i}\Gamma_{SD}} + \frac{c_4}{\Gamma_{SD}\Gamma_{RD_i}} \tag{9}$$

where Γ_{SD}, Γ_{SR_i} and Γ_{R_iD} are the average SNR of each corresponding link, and constants $c_1 \sim c_4$ are determined by spectrum efficiencies of each link which involves both channel coding rate and modulation order. Furthermore, the outage probability of *SD* link is approximated as

$$\Pr(E_{SD}) = 1 - \exp(-\frac{\gamma_{th1}}{\Gamma_{SD}}) \approx \frac{\gamma_{th1}}{\Gamma_{SD}}, \tag{10}$$

by using $e^{-x} \approx 1 - x$ when x is relatively small, corresponding to high SNR region. Hence, letting

$\Gamma = \Gamma_{SD} = \Gamma_{SR_i} = \Gamma_{R_iD}$, when SNR approaches infinity, we arrive at

$$\Pr(E_{R^*}) \approx (\frac{c_1+c_2+c_3+c_4}{\Gamma^2})^m \div (\frac{\gamma_{th1}}{\Gamma})^{m-1} = \frac{(c_1+c_2+c_3+c_4)^m}{\gamma_{th1}^{m-1}\Gamma^{m+1}}. \quad (11)$$

Consequently, from (11) the OLF relaying system is able to achieve $m+1$ order diversity gain. ■

IV. STATISTICAL LEARNING-BASED RELAY SELECTION ALGORITHM

In this section, we first provide the criterion of selecting relays for our system, and then propose a supervised learning-based relay selection algorithm.

A. RELAY SELECTION CRITERION

For the existing opportunistic DF system, i.e., intra-link error is not allowed, the selection criterion for choosing R^* is

$$R^* = \arg\max_{R_j \in \mathbb{R}'}\{\gamma_{R_jD}\}, \quad (12)$$

with γ_{R_jD} being the instantaneous channel SNR of R_j to the destination link [12]. The set $\mathbb{R}'$ represent the relay sets in which the relay node successfully decodes the source message. This is due to the reason that maximal-ratio combining (MRC) depends on the channel SNR and will take more benefit if the channel SNR is higher. Moreover, it should be noted here that [12] does not adopt MRC at the destination. However, for fair comparison, we use MRC in existing ODF relaying system to combine the signals from the source and the selected relay. However, if intra-link errors are allowed, the situation will be changed.

Based on the theoretical analyses given in [21], the whole LF system can be recognized as a model of *source coding with a helper* problem, i.e., the message transmitted by the relay which is shown in Fig. 3 acts as the side information. A truth is that if the correlation between the source and the side information gets high, the system performance becomes better. Intuitively, if we chop the source message into several pieces, then the side information with higher correlation contains more pieces of the source message. Therefore, it is much easier to decode the source message. In theory, the higher correlation results in a larger achievable rate region, which is shown in *Appendix* A. Consequently, to achieve an error probability as low as possible, as high as possible correlation is required. Moreover, in the existing LF systems, the correlation is determined both by the intra-link error probability and the *RD* link error probability, i.e., we should fairly consider the impact of both *SR* and *RD* links. Therefore, the best relay R^* for OLF system from *source coding with a helper* viewpoint is

$$R^* = \arg\min_{R_i \in \mathbb{R}}\{\alpha_i \star \beta_i\}, \quad (13)$$

where α_i and β_i are the bit error probability of SR_i and R_iD links. The binary convolution operator is denoted by $\star$, i.e., $a \star b = (1-a)b + (1-b)a$. With a smaller value on $\alpha_i \star \beta_i$, the achievable rate region becomes larger. Alternatively, the outage probability is minimized. Hence, we use the criterion of minimizing $\alpha_i \star \beta_i$ to select the relay nodes, which is deriving from the theoretical rate region of the source coding with a helper. The bit error probability α is obtained using lossy source-channel separation theorem [21]

$$\alpha = \begin{cases} H_2^{-1}[1 - \frac{\log(1+\gamma)}{R_c}] & \gamma < \gamma_{th} \\ 0 & \gamma \geq \gamma_{th}, \end{cases} \quad (14)$$

where the inverse binary entropy function[1] is denoted by $H_2^{-1}(\cdot)$, γ is the instantaneous channel SNR, and γ_{th} is the threshold SNR for spectrum efficiency R_c. A detailed explanation is given in *Appendix* A. It should be emphasized here that the estimation of the error probability may not be accurate enough if γ exceeds the threshold γ_{th}, which corresponds to the high SNR regime. The estimation can cause some performance losses on the outage probability.

Therefore, we have the selection policy (referred as theoretical probability-based selector) for the OLF relaying system of which the aim is to minimize the binary convolution of the intra-link and the *RD* link error probabilities. To make comparison, we further propose a round-robin selection (RRS) based on the automatic repeat request (ARQ) message. In detail, the destination node sends an ARQ to activate one relay node to assist the decoding in each round, until the decoding is succeed or there is no relay node left. An outage happens if and only if none of the relay nodes can assist the decoding. Definitely, this is the optimal scenario according to *Definition* 1.

B. RELAY SELECTION ALGORITHM

In fact, it may be hard to deploy the theoretical probability-based selector in the OLF relaying system. The reason is two-fold: (1) In some practical scenarios, it is impossible to obtain the error probability from the channel state information. Even it is possible, the channel may vary very fast and a lot of computational resources is required to estimate both *SR* and *RD* links. (2) The assumption behind derivation of the theoretical error probability is that the block length goes to infinity. However, in practice, the block length of messages is limited. Inspired by the machine learning, if we can find a pattern (even a hidden pattern) associated to a specific practical scenario for performing relay selection, then it is possible to implement our OLF system practically. Actually, our relay selection problem is equivalent to a classification problem in the field of machine learning, which classifies the relays into two classes, best relays and the other relays. Classification problem can be easily solved by statistical learning algorithms. For our OLF relaying system, let *i* define the label of each relay, and the classification algorithm searches the

[1] In fact, binary entropy function does not have its inverse form. However, we are only interested in the interval [0, 0.5] since the bit error probability falls into this interval. Thereby, it is able to get the inverse function in this interval where H_2^{-1} only takes value from [0, 0.5].

possible labels of the best relays, with the instantaneous SNR being the inputs.

A supervised learning algorithm, in particular, a fully connected 3-layer feedforward neural network (FNN) [29], is adopted in the OLF relaying system to find R^*. We set the numbers of nodes for input, hidden, and output layers at 5, 10, and 2, respectively, to train an FNN selector for 2 relays. The objective of finding the optimal selector Φ^* is

$$\Phi^* = \arg\max_{\Phi\in\Delta}\{\mathbf{1}_{\mathbb{R}^*}[\Phi(\mathbb{H})]\}, \tag{15}$$

where Δ denotes the set of all possible selectors, $\mathbb{H}$ denotes the set of CSI (i.e., the instantaneous SNR for each link) for all transmission samples, $\mathbb{R}^*$ denotes the set of best relays defined in Definition 1.

We adopt traversal method to obtain the training data. More specifically, we repeat the simulations for our system with 2 relays until we obtain enough data. In the simulations, the destination D performs joint decoding with each activated relay respectively. A raw training data set $\mathscr{S}_{raw}$ that contains the transmission data is defined as

$$\mathscr{S}_{raw} = \{\phi \mid \phi = \{\mathbb{R}_C, \mathbb{T}\}\}, \tag{16}$$

where $\mathbb{R}_C$ is the set of candidate relays with which an outage event does not happen, and $\mathbb{T}$ is the set of CSI of each transmission. In $\mathscr{S}_{raw}$, the data corresponding to the cases that the number of best relays is 2 or the transmission outage happens should be excluded, because it is not interesting for the relay selector. Therefore, we obtain the valid training data set as

$$\mathscr{S} = \{\phi \mid \phi \in \mathscr{S}_{raw}, |\phi[\mathbb{R}_C]| = 1\}. \tag{17}$$

It should be emphasized that this relay selection algorithm can be easily extended to multiple relay scenarios. After obtaining the trained FNN selector for 2 relays, selection for multiple relay scenarios can be performed recursively as

$$R^* = \Phi(\Phi(\Phi(R_2, R_1)\ldots, R_{m-1}), R_m). \tag{18}$$

The learning-based relay selection algorithm is summarized in **Algorithm** 1.

Algorithm 1 Lossy Forwarding Relay Selection Algorithm

Input: Instantaneous SNRs of all links.
Output: The "best" relay R^*.
/* Initialization. */
Let $R^* = R_1$;
Set m as number of relays;
Set Φ as the trained FNN selector;
for $i = 2{:}m$ **do**
  /* Run FNN selector recursively. */
  $R' = \Phi(\gamma_{sd}, \gamma_{sr^{(i)}}, \gamma_{rd^{(i)}}, \gamma_{sr^{(*)}}, \gamma_{rd^{(*)}})$;
  if R' *is not* R^* **then**
    /* Found a better relay.*/
    Let $R^* = R'$;
  end
end
return R^*;

V. NUMERICAL RESULTS AND DISCUSSIONS

A. PARAMETER SETTING

First of all, the main configuration parameters used in the following numerical investigations are described in this sub-section.

- Information length: $n = 2000$
- Interleaver type: random
- Relay number: 1, 2, and 3
- Outer code: a 1/2 rate non-recursive systematic convolutional code with $[03, 02]_8$
- Inner code: a 1/2 rate recursive systematic convolutional code with $[03, 02]_8$ and doping rate one
- Mapping: binary phase-shift keying (BPSK) and quadrature phase-shift keying (QPSK) with Gray mapping.
- Decoder: the iteration time is set at 10 between outer decoders, referred to as vertical iteration, and 30 between inner and outer decoders, for the proposed scheme. Moreover, the intra-link error P_e is assumed to be known by the receiver. MRC is used for the conventional DF scheme, with the iteration time set at 10.
- Performance metric: FER, outage probability.

B. INVESTIGATION OF APPROXIMATION ACCURACY

In this sub-section, the accuracy of our approximated theoretical derivation is investigated, by comparing with Monte Carlo simulations, as shown in Fig. 4, where two location scenarios (location A and B in Figure 2) are considered. Clearly, in the 2-relay case, the outage gap in terms of SNR is around 0.5 dB between those achieved by approximated bound and Monte Carlo approaches, respectively. The gap is found to be 0.85 dB for the 3-relay scenario. Moreover, it can be seen that the outage probabilities achieved in Location B

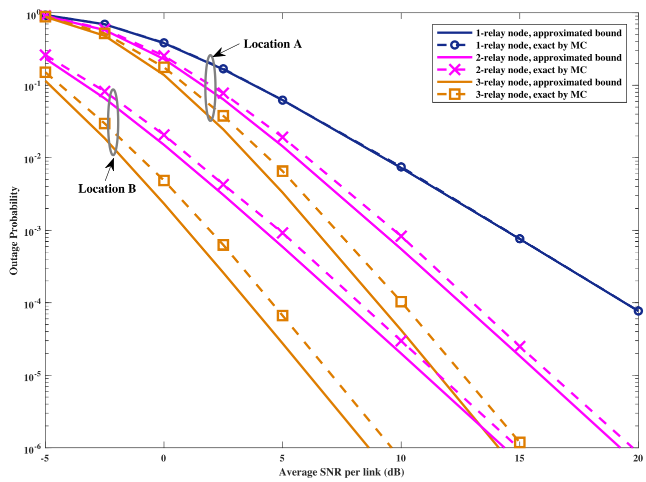

FIGURE 4. Outage comparison between Monte Carlo simulations and theoretical derivations.

are generally lower than those obtained in Location A, due to the geometric gain of SR and RD links.

C. IMPACT OF LEARNING PARAMETERS

In our simulation, 5658 and 3622 items are fed to the system, in the case of BPSK and QPSK, respectively, aiming at the proposed FNN selector training. Such training process will be conducted 2000 times. The ratio of training data by test data is 7:3, where the regularizing value λ utilized in the cost function is set at 0.5. For BPSK, both the training and test accuracy are very close to 100%, as shown in Fig. 5, while the two accuracy indexes for QPSK are also larger than 90%.

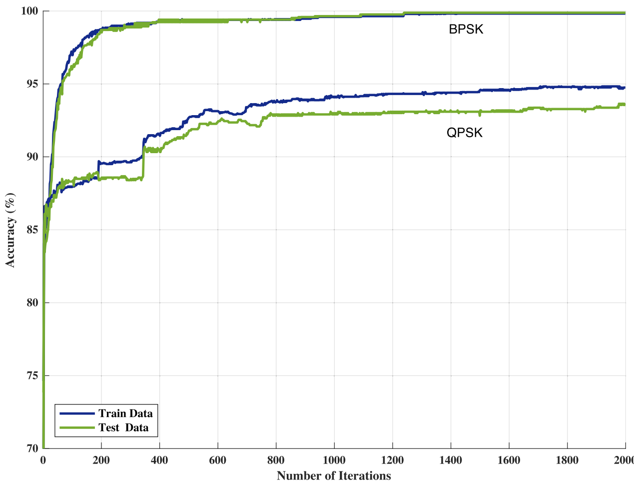

FIGURE 5. Training performance of supervised learning of 3-layer FNN for relay selection.

D. PERFORMANCE COMPARISON TO OTHER SCHEMES

In this sub-section, a series of numerical results are investigated. Specifically, the outage performances, achieved by different relaying techniques, selection algorithms, as well as the modulation schemes, are studied.

First of all, Monte Carlo simulations are conducted in the cases of 1 $\sim$ 4 relays, in order to evaluate the proposed relay selection strategy. As the counterpart technique, a RRS is presented regarded as the optimal case according to Definition 1. In RRS, relay is activated only when the ARQ message is received. Specifically, as long as the destination does not achieve successful decoding, an ARQ will be sent to randomly select a relay R_i for help, until successful decoding is achieved. If the destination can not decode correctly after trying all relays, the system outage will happen. For fair comparisons, decoding process will be conducted with the signal coming from the relay which is latest activated.

According to Fig. 6, clearly, OLF with the proposed theoretical probability-based selection strategy is found to achieved almost the same outage performances as the optimal RRS scheme, regardless of the number of the relay. The performance gain over the conventional ODF is roughly 1 dB by fixing the outage at 0.001, in terms of the SNR.

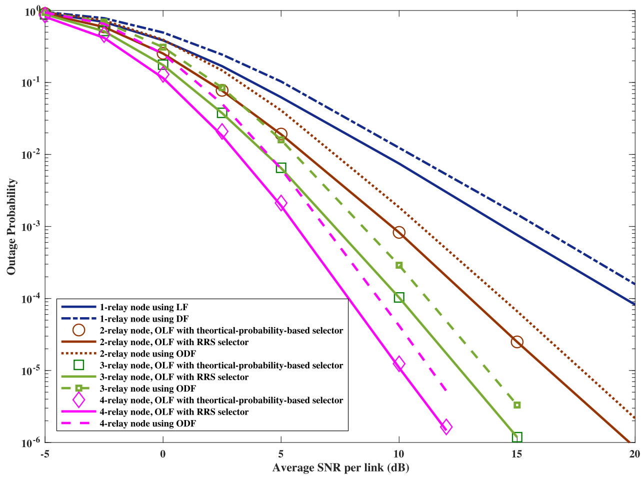

FIGURE 6. Outage probability comparisons by Monte Carlo simulations, among RRS, ODF and the proposed OLF systems.

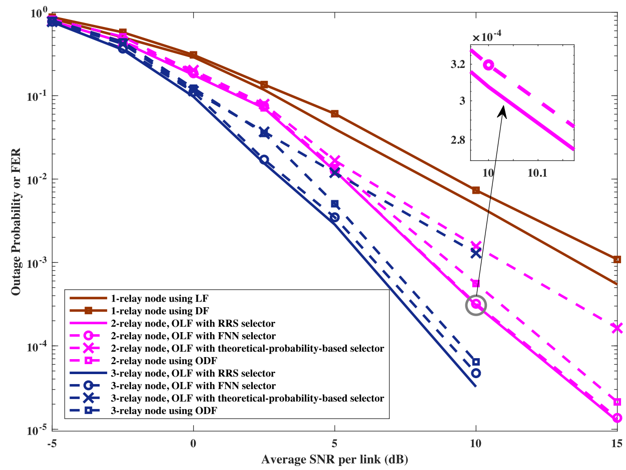

FIGURE 7. FER simulation results comparisons of different selection schemes in the case of using BPSK.

Moreover, a $m+1$ order diversity can be achieved if m relays are deployed.

Then, the FER simulation results obtained by the proposed FNN selector in OLF are presented in Fig. 7, which are also compared with the optimal RRS, as well as ODF utilizing the relay selection strategy as described in Section IV. Similar to the Monte Carlo simulations, the FER curves obtained by the FNN selector in OLF is found very close to those with the optimal RRS scheme, and the diversity order also follows the findings mentioned above. Note that, the theoretical probability-based selector work well in lower SNR region, but it leads to significant performance loss as the average SNR increases, and finally make the outage curves converge to the second order diversity for both 2-relay and 3-relay scenarios.

Moreover, the outage performances with BPSK and QPSK modulations are also compared in Fig. 8, by fixing the relay number equal to 2. Obviously, the scheme with BPSK results in much lower outage probabilities, where the relative

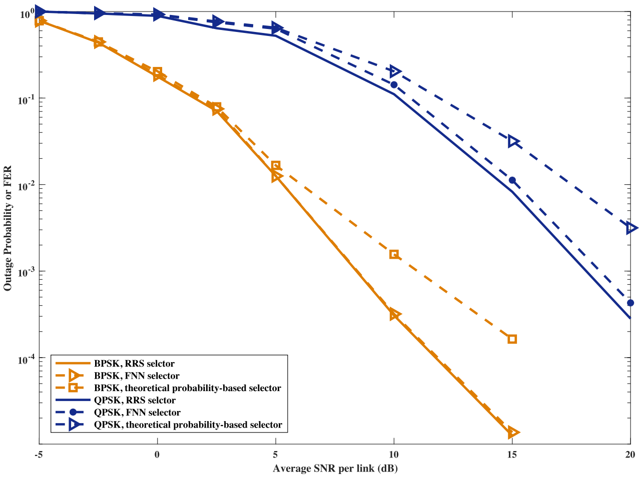

FIGURE 8. Performance comparison in terms of FER using BPSK and QPSK with different selection algorithms.

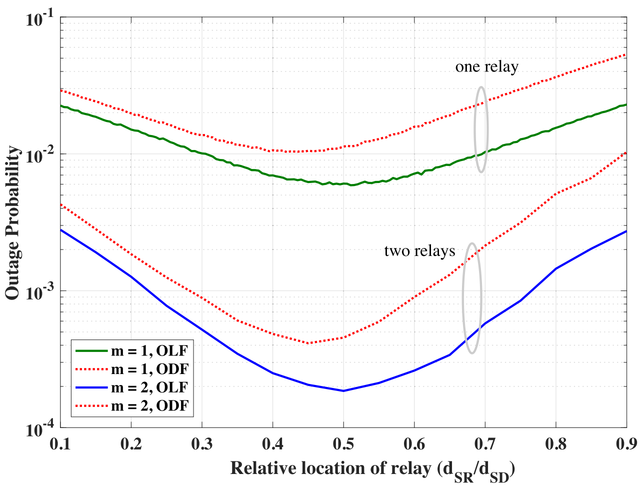

FIGURE 9. The outage probability versus the relay location.

performances among different schemes exhibit the same tendency as mentioned above.

E. IMPACT OF RELAY DEPLOYMENT

In this sub-section, the influence of relay location is investigated. Specifically, we consider a varied location of relay on the direct line between the source and destination nodes, which is described in Fig. 2. The geometric gain of SR link can be given by

$$G_{SR} = \left(\frac{d_{SD}}{d_{SR}}\right)^{\alpha}, \tag{19}$$

where $\alpha = 3.52$. Similarly, the geometric gain of RD link, denoted by G_{RD} can be also calculated. We normalize the distance of SD link to 1, and set the relay location from 0.1 to 0.9, denoting the distance ratio of SR link and SD link, i.e., d_{SR}/d_{SD}. The average SNR of *SD* link was set at 5 dB. According to Fig. 9, the best relay location, yielding the lowest outage probabilities are found about the middle of the SD link for the OLF relaying system. However, it is different for the ODF relaying system, it is better to locate the relay node a little bit closer to the source node. Moreover, the performance gain of the proposed OLF over ODF is found larger when the relay is approaching the destination. The observations are similar for both the 1-relay and 2-relay cases.

VI. DISCUSSION

Based on the numerical results, we have the following remarks.

- As the number of relay nodes increases, the gap between our approximated outage probability (8) and the exact outage probability (1) becomes large. The reason is that in the approximation, we introduce the event E_{SD} as a conditional event to independently handle the events $E_{R_i}|E_{SD}$. However, there exists correlation between the events E_{R_i} which is not fully taken into account in this work. Hence, if the number of the relay nodes increases, the approximation has more loss due to the effect of the correlation.
- At the high SNR regime, the theoretical probability-based selection has some losses due to the reason that the estimation on the error probabilities α_i and β_i is from the rate-distortion function. The estimation may have errors in practice because the rate-distortion function requires random and long enough block length. At the high SNR regime, the estimation may obtain the estimated values on α_i and β_i as 0 which makes the relays indistinguishable. In this case, it will randomly select a relay node to compute the outage probability. The RRS-based method selects the relay nodes one-by-one and choose the optimal one, while the FNN-based method adopts the generated channel realization data and evaluates the outage performance. Thus, the theoretical-based method has the losses compared to other selection schemes. The loss comes from the estimation rather than the selection criterion itself. If we have a significantly accurate estimation, then the theoretical-probability-based method can also converge to the performances of the other schemes. However, it is difficult to estimate the error probability if the coding scheme is involved.
- Regarding the effect of the relay location, the OLF system has the best outage performance if the relay node locates in the middle. The reason is that the OLF system could achieve better performance if *SR* and *RD* links have equal contributions to the outage performance, from the expression $\alpha \star \beta$. However, it is better to increase the probability of successful decoding in the ODF system which the relay can participate the cooperation with more opportunities. Thereby, the *SR* link has more contributions to the outage performance.

VII. CONCLUSION

In this paper, we have investigated the system performance of an opportunistic lossy-forwarding system. We have examined the outage probability and given an approximated theoretical bound on the outage probability. A theoretical opportunistic relay selection strategy for lossy-forwarding system based on source coding with side-information theory has been proposed. For practical system, a machine learning-based selection method has been designed in the OLF relaying system. We have conducted a series of simulations, where both theoretical results and simulation results have shown that full diversity order can be achieved by the proposed system. Numerical results have indicated that our system outperformed conventional opportunistic decode-and-forward system in both theory and practice. Finally, we have discussed the relay deployment of the OLF relaying system, where the optimal position is found to be in the middle of the *SD* link.

APPENDIX A THE THEORETICAL VIEWPOINT OF LOSSY-FORWARDING SYSTEM

Based on the existing works, it is able to model the lossy-forwarding system as a source coding with a helper problem. Here, as shown in Figure 10, the relay node acts as a helper, and the destination takes the advantage of the correlation knowledge between the source link and the helper link, for reconstructing the estimations of the source message. In the network information theory, we know that there is an achievable rate region that the source message could be recovered with an arbitrarily low error probability, i.e., if the rate pair $(R_1, R_2)^2$ falls outside the achievable rate region, it is impossible to achieve lossless communication. As shown in Figure 10, this case corresponds to an outage event. Hence, we only need to examine the area of the inadmissible region to determine the outage probability. To this end, we have the following steps to meet the goal of deriving the outage probability.

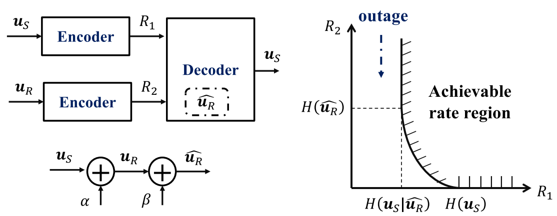

FIGURE 10. The abstract LF relaying model from the viewpoint of source coding with a helper and its achievable rate region.

Step 1 Calculating the Intra-Link Error Probability: We can theoretically calculate the intra-link error using Shannon's lossy source-channel separation theorem, as

$$H_2(\alpha) = \frac{\log_2(1+\gamma_{SR})}{R_{c0}}, \tag{20}$$

with α being the error probability and R_{c0} denoting the spectrum efficiency (joint consider the channel coding rate and the modulation scheme) of the *SR* link. The function $H_2(\cdot)$ is the binary entropy function. Therefore, we can obtain the intra-link error probability α by taking the inverse function of the binary entropy function. Obviously, the error probability varies with respect to the instantaneous SNR of the *SR* link.

Step 2 Determining the Achievable Rate Region: Based on the source coding with a helper problem, the achievable rate region is given by

$$\begin{cases} R_1 \geq H(\mathbf{u}_S|\hat{\mathbf{u}}_R) \\ R_2 \geq I(\mathbf{u}_R; \hat{\mathbf{u}}_R) \end{cases}, \tag{21}$$

where $H(\cdot|\cdot)$ and $I(\cdot;\cdot)$ represent the conditional entropy and mutual information respectively. The estimation $\hat{\mathbf{u}}_R$ is the reconstructed version of the relay message at the destination node. Now we need to calculate the conditional entropy and mutual information terms. From Figure 10, we can treat $\mathbf{u}_R$ as the bit-flipped version of $\mathbf{u}_S$ with flipping probability α, and $\mathbf{u}_R$ as the bit-flipped version of $\mathbf{u}_R$ with probability β. Then, after several elementary calculation steps, we have

$$\begin{cases} R_1 \geq H(\mathbf{u}_S|\hat{\mathbf{u}}_R) = H_2(\alpha \star \beta) \\ R_2 \geq I(\mathbf{u}_R; \hat{\mathbf{u}}_R) = 1 - H_2(\beta) \end{cases}. \tag{22}$$

Step 3 Connecting the Channel Variation and the Rate Region: Now, we turn to build the relation between the rate pair (R_1, R_2) and the channel SNR γ_{SD} and γ_{RD}. By adopting the source-channel separation theorem, we have

$$\begin{cases} R_1 \leq \dfrac{\log_2(1+\gamma_{SD})}{R_{c1}} \\ R_2 \leq \dfrac{\log_2(1+\gamma_{RD})}{R_{c2}} \end{cases}, \tag{23}$$

where γ_{SD} and γ_{RD} are the channel instantaneous SNR of the *SD* and *RD* links, and R_{c1} and R_{c2} are the spectrum efficiency of the *SD* and *RD* links, respectively.

By using these facts, we can substitute (23) and (20) into (22) with the *pdf* function of the channel instantaneous SNR γ_i. The detailed calculation is omitted since the reader can easily find it in [21]. Finally, we have

$$\Pr(E_{R_i}) = \Pr\{(R_1, R_2) \notin \text{achievable rate region}\}. \tag{24}$$

APPENDIX B EXPRESSION OF $\Pr(E_{R_i})$

Following the steps in Appendix A, we obtain the exact outage probability of the lossy-forwarding system with a single relay, i.e., $\Pr(E_{R_i})$, as [21]

$$\Pr(E_{R_i}) = \frac{1}{\Gamma_2}[\exp(-\frac{\psi^{-1}(1)}{\Gamma_1})]\int_0^{\psi^{-1}(1)} \exp[-\frac{\gamma_2}{\Gamma_2}] \times \left[1-\exp(-\frac{\psi^{-1}(1-\psi(\gamma_2))}{\Gamma_0})\right] d\gamma_2 \tag{25}$$

$$+\frac{1}{\Gamma_1}[\exp(-\frac{\psi^{-1}(1)}{\Gamma_2})]\int_0^{\psi^{-1}(1)} \exp[-\frac{\gamma_1}{\Gamma_1}] \times \left[1-\exp(-\frac{\psi^{-1}(1-\psi(\gamma_1))}{\Gamma_0})\right] d\gamma_1 \tag{26}$$

[2]The rate pair includes the source and relay source coding rates. Actually, we did not perform source coding in this work, the rate pair is an intermediate variable to build the relation between the source coding and the channel variations.

$$+\frac{1}{\Gamma_1\Gamma_2}\int_0^{\psi^{-1}(1)}\int_0^{\psi^{-1}(1)}\exp[-\frac{\gamma_1}{\Gamma_1}-\frac{\gamma_2}{\Gamma_2}] \times\left[1-\exp(-\frac{\psi^{-1}(\Psi(\gamma_1,\gamma_2))}{\Gamma_0})\right]d\gamma_0 d\gamma_2 \quad (27)$$

with $\{0, 1, 2\}$ denoting the SD, SR_i, R_iD links, and

$$\psi^{-1}(x)=\begin{cases}2^{2xR_c}-1 & \text{one-dimensional signal}\\ 2^{xR_c}-1 & \text{two-dimensional signal,}\end{cases}$$

$$\Psi(\gamma_1,\gamma_2)=H_2\{H_2^{-1}(1-\psi(\gamma_1))\star H_2^{-1}(1-\psi(\gamma_2))\},$$

$$\psi(\gamma)=\begin{cases}\frac{1}{2R_c}\log_2(1+\gamma) & \text{one-dimensional signal}\\ \frac{1}{R_c}\log_2(1+\gamma) & \text{two-dimensional signal.}\end{cases}$$

REFERENCES

[1] I. Kim and D. Kim, "Minimizing Source-Sum-Power consumption in multi-sensor single-relay networks," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2362–2366, Sep. 2011.

[2] A. Razi, F. Afghah, and A. Abedi, "Power optimized DSTBC assisted DMF relaying in wireless sensor networks with redundant super nodes," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 636–645, Feb. 2013.

[3] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[4] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[5] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3524–3536, Dec. 2006.

[6] T. Cui, F. Gao, T. Ho, and A. Nallanathan, "Distributed space-time coding for two-way wireless relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 658–671, Feb. 2009.

[7] H. A. Suraweera, M. Soysa, C. Tellambura, and H. K. Garg, "Performance analysis of partial relay selection with feedback delay," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 531–534, Jun. 2010.

[8] M. Seyfi, S. Muhaidat, and J. Liang, "Performance analysis of relay selection with feedback delay and channel estimation errors," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 67–70, Jan. 2011.

[9] Q. F. Zhou, Y. Li, F. C. M. Lau, and B. Vucetic, "Decode-and-forward two-way relaying with network coding and opportunistic relay selection," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3070–3076, Nov. 2010.

[10] T. T. Duy and H. Y. Kong, "Exact outage probability of cognitive two-way relaying scheme with opportunistic relay selection under interference constraint," *IET Commun.*, vol. 6, no. 16, pp. 2750–2759, Nov. 2012.

[11] K. Stamatiou, D. Chiarotto, F. Librino, and M. Zorzi, "Performance analysis of an opportunistic relay selection protocol for multi-hop networks," *IEEE Commun. Lett.*, vol. 16, no. 11, pp. 1752–1755, Nov. 2012.

[12] J. Hu and N. C. Beaulieu, "Performance analysis of decode-and-forward relaying with selection combining," *IEEE Commun. Lett.*, vol. 11, no. 6, pp. 489–491, Jun. 2007.

[13] Y. Han, S. H. Ting, C. K. Ho, and W. H. Chin, "Performance bounds for two-way amplify-and-forward relaying," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 432–439, Jan. 2009.

[14] J. He, V. Tervo, X. Zhou, X. He, S. Qian, M. Cheng, M. Juntti, and T. Matsumoto, "A tutorial on lossy forwarding cooperative relaying," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 66–87, 1st Quart., 2019.

[15] E. Abbe and M. Warnwright, "Tutorial, topic: Information theory meets machine learning," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2015, pp. 1–46. [Online]. Available: http://www.princeton.edu/~eabbe/publications/tuto_slides_part1.pdf

[16] B. M. Kurkoski and H. Yagi, "Quantization of binary-input discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4544–4552, Aug. 2014.

[17] F. J. C. Romero and B. M. Kurkoski, "Decoding LDPC codes with mutual information-maximizing lookup tables," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 426–430.

[18] V. Misra and T. Weissman, "Unsupervised learning and universal communication," in *Proc. IEEE Int. Symp. Inf. Theory Process. (ISIT)*, Oct. 2013, pp. 261–265.

[19] K. Anwar and T. Matsumoto, "Accumulator-assisted distributed turbo codes for relay systems exploiting source-relay correlation," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1114–1117, Jul. 2012.

[20] J. Garcia-Frias and Y. Zhao, "Near-Shannon/slepian–wolf performance for unknown correlated sources over AWGN channels," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 555–559, Apr. 2005.

[21] X. Zhou, M. Cheng, X. He, and T. Matsumoto, "Exact and approximated outage probability analyses for decode-and-forward relaying system allowing intra-link errors," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 7062–7071, Dec. 2014.

[22] M. Cheng, K. Anwar, and T. Matsumoto, "Outage probability of a relay strategy allowing intra-link errors utilizing slepian-wolf theorem," *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 1, p. 34, Feb. 2013, doi: 10.1186/1687-6180-2013-34.

[23] A. Wolf, M. Matthe, A. Festag, and G. Fettweis, "Outage based power allocation for a lossy forwarding two-relaying system," in *Proc. IEEE 20th Int. Workshop Comput. Aided Model. Design Commun. Links Netw. (CAMAD)*, Sep. 2015, pp. 267–272.

[24] A. Wolf, D. Cristina González, J. Cándido, and G. Fettweis, "Asymptotically optimal power allocation for WSNs with mutually correlated sensing data," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1317–1320, Jul. 2016.

[25] D. C. Gonzalez, A. Wolf, L. L. Mendes, J. C. S. Santos Filho, and G. P. Fettweis, "An efficient power allocation scheme for multirelay systems with lossy intra-links," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1549–1560, Apr. 2017.

[26] S. Qian, M. Cheng, and T. Matsumoto, "Outage based power allocation for a lossy-forwarding relaying system," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 2139–2144.

[27] W. Jiang, X. He, and T. Matsumoto, "Power allocation in an asymmetric wireless sensor network," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 378–381, Feb. 2017.

[28] S. Qian, J. He, M. Juntti, and T. Matsumoto, "Fading correlations for wireless cooperative communications: Diversity and coding gains," *IEEE Access*, vol. 5, pp. 8001–8016, 2017.

[29] P. Auer, H. Burgsteiner, and W. Maass, "A learning rule for very simple universal approximators consisting of a single layer of perceptrons," *Neural Netw.*, vol. 21, no. 5, pp. 786–795, Jun. 2008.

HAIYANG ZHANG received the B.Eng. degree from the School of Computer Science and Technology, Huaibei Normal University, Huaibei, China, in 2003, and the M.Eng. degree in computer technology from the School of Computer and Information, Hefei University of Technology, Hefei, China, in 2011. He is currently an Associate Professor with the School of Environment and Geomatics, Suzhou University, Suzhou, China. His current research interests mainly include cloud computing, database systems, and wireless communications.

ZHIWEI ZHANG received the M.Sc. degree in computer system architecture from the School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China, in 2013, and the Ph.D. degree in computer science and technology from the South China University of Technology, in 2016. He is currently an Associate Professor with the School of Informatics and Engineering, Suzhou University, Suzhou, China. His research interests include graph neural networks, social computing, complex networks, data science, and big data technology.