Multiple Relay Selection for Delay-Limited Applications

Ahmad Alsharoa, Member, IEEE, Walid Abediseid, Member, IEEE, and Mohamed-Slim Alouini, Fellow, IEEE

Abstract—A multiple relay selection system model that implements the decode-and-forward mode is investigated. All communication nodes are assumed to be equipped by multiple antennas. Furthermore, lattices space-time coded multiple-input multiple-output half duplex channel is applied. The main goal is to increase the throughput of the system by selecting multiple number of relays. The selection criteria depends on the maximum decoding delay at relays where the system implements a decoding time-out algorithm at each relay. This leads to a significant saving in the overall system power consumptions and attempts to solve the relays synchronization problem. All results are presented using numerical simulations.

Index Terms—Multiple-input multiple-output, lattices space-time, multiple relay selection.

I. INTRODUCTION

COOPERATIVE communication techniques have recently attracted significant attention in order to enhance both the data rate and the reliability of wireless networks. These improvements can be achieved by introducing communication nodes in the network to serve each other in various communication tasks [1].

While most of existing relay selection schemes deal with single relay selection in order to maximize the Signal-to-Noise Ratio (SNR), little attention has been given to the multiple relay selection problem. Relay selection technique is crucial in improving the performance of wireless systems. We can achieve better performance by selecting multiple relays instead of a single relay [2]. In [3], the authors studied the capacity of multiple-input multiple-output (MIMO) relay systems, and in [4] a multi-relay selection scheme for multi-stream cooperative MIMO system were investigated. It must be noted that all the aforementioned works focused on using relay selection criteria in order to maximize the SNR of the system.

Most of previous works regarding cooperative systems assumed that all relays are perfectly synchronized during transmission, which is generally very difficult to maintain in practice. It has been found that a significant performance degradation occurs and that the benefits of cooperation may even vanish if the synchronization time errors are large [5], [6]. Assuming all relays receive the source message at the same time, synchronization errors may occur at relays due to the time delay in attempting to decode the message. In [7], the authors focus on buffered relays under the single antenna case for which the source starts transmitting its next packet as soon as a relay has decoded the current packet. The packet is then queued in the relays buffer for transmission in a first-come first-serve fashion to the destination. To the best of our knowledge, the multiple relay selection problem for the Decode-and-Forward (DF) mode that implements a decoding time-out algorithm at the relays has not been yet studied.

In this paper, a multiple MIMO relay selection with DF relay mode is considered. We assume that the message is encoded using LAttices Space-Time (LAST) coding schemes. These schemes enjoy many advantages in terms of achieving high data rates, excellent error performance, and low decoding complexity. Also we assume a coded system where optimal or near-optimal decoders based on tree search algorithms (such as the well-known sphere decoder or the sequential decoder) are implemented to achieve optimal or near-optimal performance with low decoding complexity. Due to the random nature of these decoders, they exhibit varying computational times to decode a message, in contrast to the optimal maximum-likelihood (ML) decoder which exhibits a fixed decoding delay but is in general prohibitively complex. The main contributions of this letter can be summarized as follows. First, we propose a multiple relay selection scheme with a new selection criteria in order to overcome the relay synchronization problems by implementing a time-out algorithm at all relays, i.e., only relays that decode the received signal within a “predetermined” time deadline are allowed to transmit their data to the destination. For the case when all the relays can decode the received signals within the deadline, there is no need to wait until delay deadline, and all relays can transmit all their signals as soon as the last relay has accomplished decoding. We then compare the performance of the proposed scheme to the performance of the one with no time deadline in which all relays are allowed to transmit their data at any time instant.

Notations: The superscript c and superscript T denote complex quantities and transpose operator, respectively. Vect(U) denotes the vectorization of the Matrix U. E{ } denotes expectation operator. We refer to $g(x) \leq a$ as $\lim_{x \rightarrow \infty} \frac{\log(a(x))}{\log(x)} \leq a$, and $\leq$ is used similarly. $\Gamma(.)$ denotes the Gamma function. $\otimes$ denotes the Kronecker product. Z, R, and C refer to the field of integers, real, and complex numbers, respectively.

II. SYSTEM MODEL

We consider a cooperative MIMO LAST coded system consisting of a source, a destination, and L relays. We assume that there is no direct link between the source and the destination. The source, the destination, and the $i^{th}$ relay have $N_s$, $N_d$, and $M_{r,i}$ antennas, respectively. During the first time slot of data transmission, the source transmits its signal to all relays. In the second time slot, only relays that can decode
the signal within a pre-determined time constraint, denoted by \( \tau_{\text{out}} \), broadcast their signals to the destination. We assume that the channel state information is available at the receiver side only.

Let us first discuss our coding scheme. Define \( \Lambda_c \triangleq \Lambda(G) = \{ z = G \rho : z \in \mathbb{Z}^m \} \) as a lattice in \( \mathbb{R}^m \), where \( G \) is an \( m \times m \) full-rank lattice generator matrix. The Voronoi cell, \( \mathcal{V}_c(G) \), with volume \( V_c = |\text{det}(G)| \) that corresponds to the lattice point \( \rho \in \Lambda_c \) is the set of points in \( \mathbb{R}^m \) closest to \( \rho \) than to any other point \( \Lambda \in \Lambda_c \). An \( m \)-dimensional lattice code \( \mathcal{C}(\Lambda_c, M_c, R) \) is the finite subset of the lattice translate \( \Lambda_c + \mathbf{u}_0 \) inside the shaping region \( R \), i.e., \( \mathcal{C} = \{ \Lambda_c + \mathbf{u}_0 \} \cap R \), where \( R \) is a bounded measurable region of \( \mathbb{R}^m \). We consider a shaping region \( R \) that corresponds to the Voronoi cell \( \mathcal{V}_c \), with volume \( V_c \), of a sublattice \( \Lambda_s = Q \Lambda_c \) of \( \Lambda_c \) (\( Q \in \mathbb{Z}^+ \)), i.e., \( \Lambda_s \subseteq \Lambda_c \). The generated codes are called self-similar nested lattice codes with nested lattice ratio given by \( Q = (V_c/V_s)^{1/m} \).

At the source, the message is first encoded into an \( N_s \times T \) space-time code matrix \( \mathbf{X}^c \) and then transmitted to all relays where \( T \) denotes the channel uses. We assume that the all-zero codeword is transmitted. We consider an outage-limited Rayleigh fading MIMO channel. During the second phase, each relay attempts to decode the received signal within a pre-determined time constraint, say \( \tau_{\text{out}} \). Three main possible scenarios exist in our system which are illustrated in Fig.1. In Case 1, all the relays decode the received signal correctly within \( \tau_{\text{out}} \) as shown in Fig.1(a). This occurs when all received signals are located close in distance, to the transmitted lattice point, say \( \mathbf{x} \). In Case 2, only some of the relays, say \( K < L \), decode the signal correctly within \( \tau_{\text{out}} \) as shown in Fig.1(b). This occurs when the received signals at \( L - K \) relays are located close to the edge of the decoding region of \( \mathbf{x} \), \( \mathcal{V}(\mathbf{x}) \). Finally, in Case 3 (see Fig.1(c)) we have that some relays decode the received signal correctly and others perform erroneous detection but all within the allowed time constraint. In this scenario, the channel can be considered as a MIMO interference channel \(^1\).

Let us analyze the performance of the overall system. In the first phase, the complex baseband model of the received signal at the \( i \)-th relay can be written as

\[
Y^c_{s,i} = \sqrt{\frac{P}{N_s}} H^c_{s,i} X^c + W^c_{s,i},
\]

where \( Y^c_{s,i} \in \mathbb{C}^{M_r \times T} \) is the received signal matrix at the \( i \)-th relay, \( H^c_{s,i} \in \mathbb{C}^{M_r \times N_s} \) is the complex channel mapping matrix between the source and the \( i \)-th relay, and \( W^c_{s,i} \in \mathbb{C}^{M_r \times T} \) is the additive Gaussian noise matrix with elements that are independent and identically distributed (i.i.d) with zero mean and unit variance, and \( \rho \) denotes the average SNR observed at the \( i \)-th relay at each receive antenna. Here, we assume that the source and all relays share the same encoding scheme. In order to simplify the analysis, one can

1In this case, the received signal can be expressed as \( y_d = \sqrt{\frac{P}{N_t}} \sum_{i \in D} H_{i,d} z + \sqrt{\frac{P}{N_t}} \sum_{i \in \mathcal{D}} H_{i,d} x_i + w_d \), where \( D \) is defined as the set of all relays with decoding time less than \( \tau_{\text{out}} \). One way to overcome the interference is to combine it with the additive noise. The performance of the system in such a case is further analyzed in Section IV.

![Fig. 1: Three possible decoding scenarios at the relay side.](image)

show (see [8]) that the equivalent real channel model of (1) can be described as

\[
y_{s,i} = \sqrt{\frac{P}{N_s}} H_{s,i} x + w_{s,i},
\]

where \( x \in \mathbb{R}^{2N_s \times T} \) is selected from a nested lattice code \( \Lambda_c \), \( w_{s,i} \) is the real Gaussian vector with i.i.d zero-mean and unit variance elements, and the real channel mapping between the source and the \( i \)-th relay is defined by

\[
H_{s,i} = I_T \otimes \begin{bmatrix} \text{Re}(H^c_{s,i}) & -\text{Im}(H^c_{s,i}) \\ \text{Im}(H^c_{s,i}) & \text{Re}(H^c_{s,i}) \end{bmatrix}.
\]

An \( N_s \times T \) space-time coding scheme is a full-dimensional LAST code if its vectorized (real) codebook (corresponding to the channel model (2)) is a lattice code with dimension \( 2N_s \times T \). As discussed in [8], the design of space-time signals reduces to the construction of a codebook \( \mathcal{C} \subseteq \mathbb{R}^{2N_s \times T} \) with code rate \( R = \frac{1}{T} \log |\mathcal{C}| = 2N_s \log_2 Q \), satisfying the input averaging power constraint \( \frac{1}{T} \sum_{\mathbf{x} \in \mathcal{C}} |\mathbf{x}|^2 \leq N_s T \).

During the second phase, the received signal at the destination can be expressed as

\[
y_d = \sqrt{\frac{P}{N_t}} \sum_{i=1}^{L} \varepsilon_i H_{i,d} x + w_d,
\]

where \( y_d \in \mathbb{R}^{2N_t \times T} \), \( x \in \mathbb{R}^{2M_r \times T} \) is the transmitted signal from the \( i \)-th relay, \( w_d \in \mathbb{R}^{2N_t \times T} \) is a real zero-mean Gaussian random variables with i.i.d elements of variance \( 1/2 \), and \( \varepsilon_i \) is the indicator random variable which equals to 1 if the \( i \)-th relay is selected and 0 otherwise.

III. MULTIPLE RELAY SELECTION FOR DELAY-LIMITED APPLICATION

For simplicity, we assume that \( N = N_s = N_d \) and \( M = M_{r,i}, \forall i = 1, \ldots, L \) (i.e., all the antennas at the relays are equal to \( M \)). We recall the following important result in [9], [10] for a point-to-point outage-limited Rayleigh fading MIMO channel. Define the multiplexing gain, \( r \), and the diversity gain \( d \), respectively as [11]:

\[
r = \lim_{\rho \to \infty} \frac{\log R(\rho)}{\log \rho}, \quad d = \lim_{\rho \to \infty} \frac{-\log P_e(\rho)}{\log \rho}.
\]

With the aid of minimum-mean square error decision-feedback equalization (MMSE-DFE) at the decoding stage, LAST coding and lattice decoding schemes, implemented via sphere [8] and sequential [10] decoding algorithms, achieve the optimal tradeoff of the channel \( d_n^{\text{out}}(r) = (M - r)(N - r), \forall 0 \leq r \leq M \).

However, since our main concern in this paper is to attempt solving the time synchronization problem and reducing
the decoding delay that exist in such systems, selecting the appropriate algorithm to perform lattice decoding is crucial in this case. It is well-known that sequential decoding algorithms achieve near-optimal performance without suffering the high decoding complexity of the ML and the sphere decoders. Such algorithms are controlled by a decoding parameter called the bias term that is responsible for the excellent performance-complexity tradeoff achieved by those decoders (see [10] and references therein for more details). These decoders will be used here to solve the time synchronization problem while reducing the time needed to decode a message.

Note that due to the random nature of the sequential decoder, the total number of active relays during each transmission is also random and depends on the decoding delay which is also random. The total number of active relays is given by \( K = \sum_{i=1}^{L} \varepsilon_i \), where \( K \) is a random variable that follows a binomial distribution given by

\[
P(K = k) = \binom{L}{k} p^k (1-p)^{L-k},
\]

where \( p \) is the probability of selecting a relay that decode the signal within the time constraint \( \tau_{out} \).

It has been shown in [10] that a sequential decoding scheme which implements a time-out algorithm can achieve the optimal DMT of the channel as long as the decoding time-out parameter exceeds a certain limit, say \( \tau'_{out} \). This is equivalent to say that

\[
\Pr(\tau_{out} > \tau'_{out}) \approx \rho^{-d_{out}(r)}, \quad \text{as } \rho \to \infty,
\]

where it has been shown that \( \tau'_{out} \) behaves as

\[
\tau'_{out} \approx 2MT + \frac{2MT}{\log(\frac{Q}{2})} \sum_{k=1}^{2MT} \frac{4\rho^k}{\Gamma(k/2 + 1)} \frac{NT(1 + \log \rho)^{k/2}}{\det(R_{kk}^T R_{kk})^{1/2}},
\]

where \( R_{kk} \) is the lower \( k \times k \) part of \( R = Q^T B_{s,i} G \). \( B_{s,i} \) is the backward filter matrix resulting from applying the MMSE-DFE to the channel gain matrix \( H_{s,i} \). \( Q \) is an orthogonal matrix corresponds to the QR decomposition of \( B_{s,i} G = QR \) (see [10] for more details). Unfortunately, equations (5) and (6) are very difficult to evaluate for low-to-moderate SNRs and hence we resort in this paper to the high SNR analysis of such a decoder. In other words, if the channel is well-conditioned, we expect to have large values of \( \det(R_{kk}^T R_{kk}) \) which allows us to select a small value for the time constant \( \tau_{out} \). On the other hand, if the channel is ill-conditioned (\( H^c \) is near singular), then \( \tau_{out} \) must take large values accordingly. However, as will be discussed in the sequel, we can choose moderate values of \( \tau_{out} \) while achieving near optimal performance.

Let \( A_1 \) represents the event that none of the relays are active and \( A_2 \) represents the decoding error event at the destination. In this case

\[
\mathcal{P}(A_1) = \Pr\left( \bigcap_{i=1}^{L} \{ \tau_{out}^{(i)} > \tau'_{out} \} \right) = \prod_{i=1}^{L} \Pr(\tau_{out}^{(i)} > \tau'_{out}) \approx \rho^{-Ld_{out}(r)}.
\]

Following the footsteps of [9], one can show that for a fixed \( K \), the asymptotic average sequential decoding error probability at the destination, averaged over the channel statistics, is given by

\[
\mathcal{P}(A_2 | K = k) = \rho^{-kd_{out}(r)} \quad \text{for } k = 1, \ldots, L.
\]

Therefore, the overall average decoding error probability is given by

\[
P_e (\rho, r) = \mathcal{P}(A_1) + \mathbb{E}_K \{ \mathcal{P}(A_2 | K) \} \approx \rho^{-Ld_{out}(r)} + \sum_{k=1}^{L} k p^{-kd_{out}(r)} = \rho^{-d_{out}(r)}.
\]

It is clear from the above analysis that there is no improvement in the achievable DMT of the proposed system results, where the system achieve the same DMT as in the case of point-to-point outage-limited MIMO channels. However, the improvement appears in the complexity of the system in terms of: 1. overcoming the synchronization problem by implementing the time-out algorithm at the decoding stage which allows all successful relays to transmit at the same time, and 2. reducing the time delay caused by the decoder at each relay trying to decode a very noisy signal. As will be shown in the sequel, our simulations show a surprising result that the proposed scheme achieves the same error performance for all SNRs compared to the same system but with no decoding time-out algorithm, i.e., a system with large decoding delay.

It must be noted that, at high SNR, on average we have all relays are active, i.e., all relays decode the source message within the time constraint. This can be seen from the following inequality on the average number of selected relays in our scheme which is given by

\[
\mathbb{E}\{ K \} = Lp = L \Pr(\tau_{out}^{(i)} < \tau'_{out}) = L \left( 1 - \rho^{-d_{out}(r)} \right).
\]

As will be shown in the sequel, this causes a great improvement in the error performance of the overall system while at the same time reducing the overall system delay.

IV. SIMULATION RESULTS AND CONCLUDING REMARKS

In this section, numerical results are provided for Rayleigh fading channels with \( M = N = 2, \ r = 0, \) and \( T = 3 \). The rate per channel use is given by \( R = 2N \log_2 Q \) (where \( Q \) is as defined in Section-II). The results have been averaged over 10\(^6\) iterations.

Fig. 2 shows the performance of the MMSE-DFE lattice sequential decoder at the destination for different values of \( L \) and \( \tau_{out} \) selected using trial and error. In particular, we select an initial value for \( \tau_{out} > 2MT \) and increment its value for the same repeated experiment. The optimal value of \( \tau_{out} \) selected in this case can achieve the full DMT with less delay. It must be noted that when \( \tau_{out} = \infty \) means that all relays decode the noisy source signal with no decoding deadline. The Frame Error Rate (FER) versus the SNR values is shown in Fig.2-a. It is depicted that the diversity order is the same for all values of \( L \) as expected from (9). In Fig. 2-b, we plot the total average delay required by the sequential decoder at the destination versus SNR, one can see that, we can divide this figure into three regions, high average delay region at the middle and two low average delay regions at the boundaries, as follows. The first region corresponds to the
region when the SNR $\leq 10$ dB and in this case we have a low average delay but a high probability of error. This is because for low SNR values the decoded point at the destination has a high possibility to be far away from the correct point. As such, it is decoded fast but with high probability of error. The second region corresponds to the 10 dB to 20 dB SNR region and has typically a high average delay. This can be explained by noting that the received point at the destination is near the cells boundary and requires as such a higher delay to decode the signal. Finally, the last region is when the SNR $\geq 20$ and correspond to the low average delay region with low probability of error since the received point become closer to the correct point as the SNR increases. Therefore, the decoding average delay and the probability of error are decreased. For instance, for $L = 8$, SNR=15 dB, and $\tau_{out} = 25$ delay unit our proposed scheme is able to achieve the same performance as when $\tau_{out} = \infty$ delay unit with about 80% saving in delay with an average delay of 15 delay unit instead of 27 delay unit.

The average number of active relays against SNR for $Q = 2$, and different values of $\tau_{out}$ is illustrated in Fig. 3. Remark that according to the higher average delay in the middle region some of relays fail to decode the signal within the time constraint. This noticed clearly, when the number of the relays in the system is large, i.e., some of the relays are selected and some are not. On the other hand, at the boundary regions almost all the relays are active.

![Fig. 2: The performance at the system for $Q = 2$ with different values of $L$ and $\tau_{out}$ against SNR (a) FER (b) Average delay.](image1)

![Fig. 3: Average number of active relays versus SNR for $L = \{4, 8, 12\}$, $Q = 2$ with different values of $\tau_{out}$.](image2)

![Fig. 4: The performance at the system for different values $Q$ versus SNR: (a) FER and (b) average delay.](image3)

To further clarify the effect of increasing the rate per channel use on the delay saving, Fig.4 plots the performance of the proposed scheme for $\tau_{out} = \{50, \infty\}$ delay unit with $L = 4$. For instance, when $Q = 2$ and SNR=17 dB, the proposed scheme saves delay about 47% since the average delay decreases from 22 delay unit to 15 delay unit by having $\tau_{out} = 50$ delay unit instead of $\tau_{out} = \infty$ delay unit. Finally, the benefits of using delay constraint appears clearly with a considerable delay saving when the rate increases. When $Q = 4$, SNR=30 dB, and using $\tau_{out} = 50$ delay unit instead of $\tau_{out} = \infty$ delay unit, our proposed scheme saves delay about 146% since the average delay decreases from 32 delay unit to 13 delay unit.

**REFERENCES**


