Nomikos, Nikolaos; Charalambous, Themistoklis; Wichman, Risto

Bandit-Based Power Control in Full-Duplex Cooperative Relay Networks

Published in:
ICC 2021 - IEEE International Conference on Communications, Proceedings

DOI:
10.1109/ICC42927.2021.9501010

Published: 01/06/2021

Document Version
Peer reviewed version

Please cite the original version:
Bandit-Based Power Control in Full-Duplex Cooperative Relay Networks

Nikolaos Nomikos
IRIDA Research Centre for Communication Technologies,
University of Cyprus,
Nicosia, Cyprus
Email: nomikos.nikolaos@ucy.ac.cy

Themistoklis Charalambous
and Risto Wichman
School of Electrical Engineering,
Aalto University,
Espoo, Finland
Emails: name.surname@aalto.fi

Abstract—Full-duplex relaying is an enabling technique of sixth generation (6G) mobile networks, promising tremendous rate and spectral efficiency gains. In order to improve the performance of full-duplex communications, power control is a viable way of avoiding excessive loop interference at the relay. Unfortunately, power control requires channel state information of both source-relay and relay-destination channels, as well as of the loop interference channel, thus resulting in increased overheads. Aiming to offer a low-complexity alternative for power control in full-duplex relay networks, we adopt reward-based learning in the sense of multi-armed bandits. More specifically, we provide bandit-based power control algorithms, relying on acknowledgements/negative-acknowledgements observations by the relay. The proposed algorithms avoid the need for channel state information acquisition and exchange, and can be employed in a distributed manner. Performance evaluation results in terms of outage probability, average throughput and accumulated regret over time highlight an interesting performance-complexity trade-off compared to optimal power control with full channel knowledge and significant performance gains over the cases without power control and random power level selection.

Index Terms—Full-duplex relaying, power control, machine learning, multi-armed bandits.

I. INTRODUCTION

Sixth generation (6G) mobile networks are envisioned to support dense deployments of small cells where users and machines will compete for wireless access. As a consequence, the available radio spectrum gets ever more crowded and, therefore, efficient spectral usage is necessitated. Towards combating this problem, novel approaches departing from the orthogonal allocation of resources in time and frequency are being developed: for example, interference coordination/mitigation mechanisms with multiple-input, multiple-output (MIMO) multi-antenna transceiver technologies. Furthermore, tremendous rate gains are expected through the use of full-duplex (FD) transceivers, offering simultaneous transmission and reception on the same spectral and temporal resources. Various antenna solutions and digital loop interference (LI) cancellation algorithms have shown the feasibility of FD relay communication with low-cost deployments in mobile networks [1], [2].

In such complex scenarios with dense deployments, a significant amount of signaling and feedback messages will be necessitated for efficient operation of the network, thus threatening the network’s performance when centralized solutions are adopted [3]. Meanwhile, in recent years, the coupling of machine learning techniques and wireless communications has shown promising results for providing low-complexity coordination mechanisms (see, for example, [4]–[6] and references therein).

A popular method of enabling FD relay operation is related to the design of efficient power control mechanisms towards avoiding excessive LI and mitigating its malicious effect on the end-to-end rate. Riikonen et al. [7] presented opportunistic relay mode selection, switching between half-duplex (HD) and FD relaying. By exploiting instantaneous CSI knowledge at the relay, transmit power adaptation maximized the instantaneous and average spectral efficiency in the uplink and downlink. In MIMO FD relay networks, Suraweera et al. [8] highlighted the performance gains of power allocation and transmit antenna selection under various cases of CSI availability. Through a simple power allocation mechanism, the zero diversity effect of using fixed power was surpassed. Then, Tran et al. [9] investigated optimal power allocation for improving the diversity of amplify-and-forward (AF) FD relaying. The closed-form expression of the derivative of the pairwise-error probability was derived and bisection was used to find the optimal power allocation, assuming that the relay had statistical knowledge of the source-relay ($S\rightarrow R$) channel, while the destination had full knowledge of the relay-destination ($R\rightarrow D$) channel. Finally, when FD relays were equipped with buffers, statistical and instantaneous CSI availability was exploited to conduct power adaptation at both the source and the selected relay, in order to maximize the chances of LI cancellation or avoidance and improve the average throughput of the network [10], [11].

Inspired by the increased density of forthcoming 6G networks and the CSI overheads of conventional power control, we aim at developing a low-complexity power control mechanism for FD relay networks. Towards this end, we invoke machine learning and more specifically, multi-armed bandits (MAB), an important area of reward-based learning algorithms. The MAB framework is not new in the context of wireless communications; see, for example, [12] in which it is reported that the MAB framework was adopted in several

Abstract—Full-duplex relaying is an enabling technique of sixth generation (6G) mobile networks, promising tremendous rate and spectral efficiency gains. In order to improve the performance of full-duplex communications, power control is a viable way of avoiding excessive loop interference at the relay. Unfortunately, power control requires channel state information of both source-relay and relay-destination channels, as well as of the loop interference channel, thus resulting in increased overheads. Aiming to offer a low-complexity alternative for power control in full-duplex relay networks, we adopt reward-based learning in the sense of multi-armed bandits. More specifically, we provide bandit-based power control algorithms, relying on acknowledgements/negative-acknowledgements observations by the relay. The proposed algorithms avoid the need for channel state information acquisition and exchange, and can be employed in a distributed manner. Performance evaluation results in terms of outage probability, average throughput and accumulated regret over time highlight an interesting performance-complexity trade-off compared to optimal power control with full channel knowledge and significant performance gains over the cases without power control and random power level selection.

Index Terms—Full-duplex relaying, power control, machine learning, multi-armed bandits.

I. INTRODUCTION

Sixth generation (6G) mobile networks are envisioned to support dense deployments of small cells where users and machines will compete for wireless access. As a consequence, the available radio spectrum gets ever more crowded and, therefore, efficient spectral usage is necessitated. Towards combating this problem, novel approaches departing from the orthogonal allocation of resources in time and frequency are being developed: for example, interference coordination/mitigation mechanisms with multiple-input, multiple-output (MIMO) multi-antenna transceiver technologies. Furthermore, tremendous rate gains are expected through the use of full-duplex (FD) transceivers, offering simultaneous transmission and reception on the same spectral and temporal resources. Various antenna solutions and digital loop interference (LI) cancellation algorithms have shown the feasibility of FD relay communication with low-cost deployments in mobile networks [1], [2].

In such complex scenarios with dense deployments, a significant amount of signaling and feedback messages will be necessitated for efficient operation of the network, thus threatening the network’s performance when centralized solutions are adopted [3]. Meanwhile, in recent years, the coupling of machine learning techniques and wireless communications has shown promising results for providing low-complexity coordination mechanisms (see, for example, [4]–[6] and references therein).

A popular method of enabling FD relay operation is related to the design of efficient power control mechanisms towards avoiding excessive LI and mitigating its malicious effect on the end-to-end rate. Riikonen et al. [7] presented opportunistic relay mode selection, switching between half-duplex (HD) and FD relaying. By exploiting instantaneous CSI knowledge at the relay, transmit power adaptation maximized the instantaneous and average spectral efficiency in the uplink and downlink. In MIMO FD relay networks, Suraweera et al. [8] highlighted the performance gains of power allocation and transmit antenna selection under various cases of CSI availability. Through a simple power allocation mechanism, the zero diversity effect of using fixed power was surpassed. Then, Tran et al. [9] investigated optimal power allocation for improving the diversity of amplify-and-forward (AF) FD relaying. The closed-form expression of the derivative of the pairwise-error probability was derived and bisection was used to find the optimal power allocation, assuming that the relay had statistical knowledge of the source-relay ($S\rightarrow R$) channel, while the destination had full knowledge of the relay-destination ($R\rightarrow D$) channel. Finally, when FD relays were equipped with buffers, statistical and instantaneous CSI availability was exploited to conduct power adaptation at both the source and the selected relay, in order to maximize the chances of LI cancellation or avoidance and improve the average throughput of the network [10], [11].

Inspired by the increased density of forthcoming 6G networks and the CSI overheads of conventional power control, we aim at developing a low-complexity power control mechanism for FD relay networks. Towards this end, we invoke machine learning and more specifically, multi-armed bandits (MAB), an important area of reward-based learning algorithms. The MAB framework is not new in the context of wireless communications; see, for example, [12] in which it is reported that the MAB framework was adopted in several.
5G cases aiming to overcome the complexity of network coordination through learning. Nevertheless, to the best of the authors knowledge, its use has not been investigated before for the problem of power control in FD relay networks.

More specifically, an online policy for selecting the transmit power in each time-slot is developed, modeling the power level selection as a MAB game. Thus, in each time-slot, the relay observes the ACK/NACK messages from the destination for the previous transmissions, as well as whether or not the receptions from the source were successful. After an initial exploration phase, the power level offering the maximum end-to-end throughput is exploited for minimizing the regret against optimal power control when full CSI is available. Our contributions are the following.

- A bandit-based power control (BB-PC) algorithm is proposed, relying on local observation by the relay of the outcome of the source’s signal reception and ACK/NACK feedback from the destination, such that the relay does not need to perform channel estimation.
- Different versions of BB-PC, based on various upper confidence bound (UCB) policies are evaluated and results are given in terms of outage probability, average throughput and accumulated regret against optimal power control with full CSI and the cases without power control and random power level selection.

The remainder of this paper is organized as follows. In Section II, we introduce the system model. In Section III, we provide in detail the MAB modeling of the power control process. The proposed bandit-based power control for FD relay networks is described in Section IV, while performance evaluation is provided in Section V. Finally, conclusions and future directions are given in Section VI.

II. SYSTEM MODEL

A two-hop cooperative network, comprising a source node $S$, a single destination $D$, and a single FD decode-and-forward (DF) relay $R$, is considered. The relay is equipped with two antennas and operates in FD mode, resulting in simultaneous transmission and reception of signals. It is considered that direct transmissions from the source towards the destination are not possible due to severe fading conditions and communication can be performed only through the relay. Fig. 1 shows an instance of the two-hop FD cooperative relay network. This simple setup is emblematic of a wide range of wireless communication applications.

Time is assumed to be divided in time-slots where source node $S$ and relay $R$ transmit using power levels $P_S$ and $P_R$, respectively. A saturated source is assumed, having always data for transmission, while the information rate is equal to $r_0$. Retransmissions rely on an ACK/NACK mechanism, where the receivers (either the activated relay or the destination) broadcast short-length error-free packets via a separate narrow-band link, informing the network on whether or not the packet transmission was successful. Furthermore, it is assumed that the wireless channel quality is degraded by additive white Gaussian noise (AWGN) and frequency flat Rayleigh block fading, following a complex Gaussian distribution characterized by zero mean and variance $\sigma_{ij}^2$ for the $i$ to $j$ link. For simplicity, the power of the AWGN is assumed to be normalized with zero mean and unit variance. Also, the channel gains $g_{ij} \triangleq |h_{ij}|^2$ are exponentially distributed [13, Appendix A]. Also, FD relaying is supported, leading to LI and $h_{RR}$ denotes the instantaneous residual LI between the two antennas of relay $R$, following a complex Gaussian distribution and taking values in the range $(0, \sigma_{RR}^2)$.

Since the relay operates in FD mode, the HD loss of conventional relays is surpassed and the destination receives one packet in each time-slot. Nonetheless, FD operation introduces LI and the power control algorithm must take into consideration the interference level arising by each transmit power level. In an arbitrary time-slot, a packet is successfully forwarded from relay $R$ towards the destination $D$ if the signal-to-noise Ratio (SNR), denoted by $\text{SNR}_{RD}$, is greater than or equal to a threshold $\gamma_D$, called the capture ratio, i.e.,

$$\frac{g_{RD} P_R}{n_D} \geq \gamma_D,$$  \hspace{1cm} (1)

where $n_D$ denotes thermal noise variance at the destination, which is considered to be AWGN.

A packet transmission from source $S$ to relay $R$ is successful if the SINR at the relay, denoted by $\text{SINR}_{SR}$ is greater than or equal to $\gamma_R$, i.e.,

$$\frac{g_{SR} P_S}{g_{RR} P_R + n_R} \geq \gamma_R.$$  \hspace{1cm} (2)

III. MAB MODELING

A. The MAB Problem

MAB refers to a class of sequential decision problems of resource allocation among several competing entities in unknown environments with an exploration-exploitation trade-off, i.e., searching for a balance between exploring all possible decisions to learn their reward distributions while selecting the best decision more often to acquire more reward. For a detailed discussion on the topic, see, for example, [14], [15]. In the classical stochastic MAB problem, introduced by Robbins [16], a player has access to a finite set of arms, and to each arm, a probability distribution with an initially unknown mean $q_j$ is associated. At each round $t$, the player chooses an arm $j$ and receives a random reward $U_{j,t}$. In our setup, the relay transmits with a power from a set of discrete power levels, $P_R$. The number of power levels $|P_R|$ and their values depend on the radio configuration. Therefore, in the MAB framework, each arm corresponds to one of the $|P_R|$ power levels.
The goal of the learner is to maximize the expected accumulated reward in the course of her interaction. If the reward distributions were known, this goal would have been achieved by always selecting the arm with highest mean reward. To identify the optimal arm, the learner has to play various arms so as to learn their reward distributions (exploration) while ensuring that the gathered knowledge on reward distributions is exploited so that arms with higher expected rewards are preferred (exploitation). The performance of the learner in implementing such an exploration-exploitation trade-off is measured through the notion of regret, which compares the cumulative reward of the learner to that achieved by always selecting the optimal arm. It is defined as the difference between the reward achieved when the best arm is pulled and the player’s choice. For our setup, the objective is to identify a policy over a finite time horizon $T$ that maximizes the expected number of successfully transmitted packets or simply, what is uniformly good adaptive learning algorithm with time horizon $T$, the number of plays of arm $i$ up to round $t$, i.e., $n_{j,t} := \sum_{s=1}^{t} I_{(l,s)=j}$, where $I_A$ denotes the indicator function of event $A$. We let $\bar{q}_{j,t}$ represent the empirical average reward of arm $j$ built using the observations from $j$ up to $t$:

$$\bar{q}_{j,t} = \frac{1}{n_{j,t}} \sum_{s=1}^{t} r_{j,s} I_{(I_{s}=j)} ,$$ (5)

where $r_{j,s}$ is the reward of arm $j$ at round $t$.

An optimistic algorithm $\pi$ maintains an index function $q_{j,t}$ for each arm $j$, which depends only on the past observations of $j$ only (e.g., $\bar{q}_{j,t}, n_{j,t}$, etc.), and that $q_{j,t} \geq q_j$ with high probability for all $t \geq 1$. Then, $\pi$ simply consists in selecting the arm with the largest index $q_{j,t}$ at each round $t$:

$$I_t = \arg \max_{j \in \mathcal{L}} q_{j,t} .$$ (6)

In the sequel, we briefly introduce some popular index policies for stochastic MABs. In the rest of this section, we assume that the reward realizations of arm $j$ belong to the interval $[0, 1]$ almost surely.

1) UCB1 [18]: UCB1 is an index policy designed based on Hoeffding’s concentration inequality for bounded random variables. The UCB1 index (or for short, UCB) is defined as follows:

$$q_{j,t}^{\text{UCB}} = \bar{q}_{j,t} + \frac{3 \log(t)}{2 n_{j,t}} .$$ (7)

2) KL-UCB [19]: KL-UCB is a novel concentration inequality for bounded random variables, and relies on the following index:

$$q_{j,t}^{\text{KL-UCB}} = \sup \left\{ \lambda \in [\bar{q}_{j,t}, 1] : \text{KL}(\bar{q}_{j,t}, \lambda) \leq \frac{\log(t) + 3 \log(\log(t))}{n_{j,t}} \right\} ,$$

where $\text{KL}(x, y)$ is the Kullback-Leibler divergence between two Bernoulli distributions with means $x$ and $y$: $\text{KL}(x, y) := x \log \left(\frac{x}{y}\right) + (1 - x) \log \left(\frac{1-x}{1-y}\right)$. When the reward distribution of arms are Bernoulli distributions, KL-UCB achieves the problem-dependent lower bound (4), and is hence said to be asymptotically optimal\(^2\). We remark that computing $q_{j,t}^{\text{KL-UCB}}$ corresponds to finding the roots of a strictly convex and increasing function\(^3\). Therefore, $q_{j,t}^{\text{KL-UCB}}$ can be computed using simple line search methods, such as bisection.

3) KL-UCB++ [20]: KL-UCB++ is a modified variant of KL-UCB, which enjoys both asymptotic and minimax optimality in stochastic MABs simultaneously. It relies on the following index:

$$q_{j,t}^{\text{KL-UCB++}} = \sup \left\{ \lambda \in [\bar{q}_{j,t}, 1] : \text{KL}(\bar{q}_{j,t}, \lambda) \leq g(n_{j,t})/n_{j,t} \right\} ,$$ (8)

\(^1\)An algorithm $\pi$ is uniformly good if for any sub-optimal arm $i$, the number of times arm $i$ is selected up to round $t$, $n_{i,t}$, satisfies: $\mathbb{E}[n_{i,t}] = o(t^\alpha)$, for all $\alpha > 0$.

\(^2\)Indeed KL-UCB is shown to be asymptotically optimal for a wider class of MABs whose reward distributions are taken within one-parameter exponential families, provided that one replaces the Kullback-Leibler divergence of Bernoulli distributions with an appropriate divergence.

\(^3\)Note that $v \mapsto \text{KL}(u, v)$ is strictly convex and increasing for $v \geq u$. 

where
\[
g(n_{j,t}) = \log_+ \left( \frac{t}{M n_{j,t}} \left( \log_+ \left( \frac{t}{M n_{j,t}} \right) + 1 \right) \right),
\]
with \( \log_+ (x) = \max(\log(x), 0) \).

## IV. Bandit-Based Power Control

### A. Online Learning Model

We now turn to model the power control problem as a MAB. Each power level corresponds to an arm, and pulling an arm corresponds to a packet transmission using the selected power level. More formally, if power level \( j \) is selected in time-slot \( t \), a reward \( r_{j,t} \) is obtained, where
\[
r_{j,t} = \begin{cases} 1, & \text{if packet received successfully,} \\ 0, & \text{otherwise.} \end{cases}
\]

Hence, the sequence \( (r_{j,t})_{t \geq 1} \) of rewards of power level \( j \) follows a Bernoulli distribution, whose mean corresponds to the probability of successful transmission using \( j \).

We consider a scenario with stationary success probabilities. In this case, success probabilities of various power levels are assumed to be fixed but unknown. Hence, for each \( j \), \( (r_{j,t})_{t \geq 1} \) is a sequence of i.i.d. Bernoulli random variables with \( \mathbb{E}[r_{j,t} | \mathcal{F}_{t-1}] = q_j \) for all \( t \), where \( \mathcal{F}_{t-1} \) denotes the set of power levels chosen by the algorithm before round \( t \), and their realized rewards.

### B. Online Learning Algorithm

We are now ready to describe our learning algorithm. After the initial exploration phase, the power level providing the maximum reward, in terms of end-to-end throughput, is exploited for minimizing the regret.

The power level selection yields a random reward from an unknown joint probability distribution, which corresponds to the links (i.e., links \{\( S \rightarrow R_j \), \( R \rightarrow D \), and \( R_j \rightarrow D \)}). In other words, pulling arm \( j \) at round \( t \) corresponds to an end-to-end packet transmission via relay \( R \). If the packet is successfully received by \( D \), a reward \( r_{j,t} \) of 1 is obtained. If an outage occurs, no reward is obtained. BB - 0RS based on distributed power control is given in Algorithm 1.

### Algorithm 1 Bandit-based power control (BB - PC) mechanism

**Input:** Set of power levels \( P_R \), capture ratios \( \gamma_R \) and \( \gamma_D \).

**for** \( t = 1, 2, \ldots \) **do**

1. Compute \( \hat{q}_{j,t} \) (5) and then \( \hat{q}_{j,t} \) according to UCB used select power level \( j \) for transmission at time-slot \( t \) using (6)
2. Receive packet from \( S \) and transmit it to \( D \)
3. if reception and transmission are successful then
   - \( r_{j,t} = 1 \)

**end**

### V. Performance Evaluation

In this section, the performance of BB – PC is evaluated, in terms of outage probability, average throughput and total accumulated regret over time. Two different BB – PC versions, based on UCB1 [18] and kl-UCB++ [20] are compared against optimal power control with full CSI (opt), no power control (no-PC) and random power level selection (rnd). The game horizon comprises \( 10^4 \) transmissions and 20 trials are conducted over which, the results are averaged. In each link, the transmit SNR ranges from 0 dB to 40 dB and it represents the ratio of the maximum available transmit power at each transmitter, assuming \( P_{S,RX} = P_{R,RS} = P_{R,MAX} \) over the noise power. Furthermore, in the outage and throughput results, the \( x \)-axis corresponds to the transmit SNR in the \( \{ R \rightarrow D \} \) link. Moreover, a fixed transmission rate \( r_0 = 3 \) bps/Hz is considered in a topology with a single relay node being able to select among six different transmit power levels, i.e., \( P_1 = 0.01 P_{MAX}, P_2 = 0.05 P_{MAX}, P_3 = 0.20 P_{MAX}, P_4 = 0.30 P_{MAX}, P_5 = 0.50 P_{MAX}, P_6 = P_{MAX} \) [21]. Regarding the wireless setting, stationary stochastic bandits where channel statistics remain the same for the whole transmission duration are assumed. Also, i.i.d. \{\( S \rightarrow R \) and \( R \rightarrow D \)\} channels with average channel SNR \( \bar{\gamma}_{S \rightarrow R} = \bar{\gamma}_{R \rightarrow D} = 0 \) dB are considered, as well as two LI cases with average channel SNR \( \bar{\gamma}_L = -10 \) dB and \( \bar{\gamma}_L = -30 \) dB.

The first outage comparison is presented in Fig. 2 for various power control algorithms when \( \bar{\gamma}_L = -10 \) dB. Here, the necessity for power control in FD communications is clearly shown, as the case with fixed transmit power at the relay (no-PC) experiences an outage floor after 20 dB. Moreover, the case of random power selection is not able to provide a satisfactory performance throughout the SNR range. On the contrary, the proposed BB – PC algorithms offer improved outage performance without an outage floor, whereas optimal power control with full CSI exhibits the best outage performance at the cost of increased overheads.
Fig. 3. Outage probability comparisons for different power control algorithms and $\gamma_{LI} = -30$ dB.

Fig. 3 depicts the outage probability for different power control algorithms when a significantly weaker LI channel is considered, characterized by $\gamma_{LI} = -30$ dB. It is clear that when full CSI is available, the optimal power control algorithm has the best outage performance. However, the two BB – PC versions (UCB1 and kl-UCB++) offer significant improvement with respect to the case without power control (no-PC) and random power level selection (rnd). Furthermore, kl-UCB++ has a clear advantage after 28 dB over UCB1 notably larger than that of the case when $\gamma_{LI} = -10$ dB. It must be outlined that both BB – PC versions avoid the outage floor that is observed when no power control is performed. Also, the two BB – PC algorithms performance surpasses that of the case where transmit power is randomly selected.

Then, average throughput results for various algorithms are illustrated in Fig. 4 when $\gamma_{LI} = -10$ dB. In this case, the LI channel power is not negligible and thus, power control is vital to maintain adequate throughput performance. More specifically, the algorithm with fixed transmit power (no-PC) has by far, the worst performance among the considered algorithms. Better results are observed through random power selection, while significantly improved performance is provided by the two BB – PC algorithms. As a result, an interesting trade-off arises, since BB – PC avoids CSI acquisition and exchange while still closely following the performance of the optimal power control algorithm (which introduces increased overheads).

The average throughput performance for different power control algorithms and a weak LI channel characterized by $\gamma_{LI} = -30$ dB is shown in Fig. 5. It can be seen that in this case, the optimal power control algorithm achieves the throughput upper bound, being closely followed by the two BB – PC algorithms. From the two BB – PC algorithms, kl-UCB++ offers a noticeable performance gain after 28 dB. Meanwhile, when power control is not employed, significantly worse average throughput is obtained. Furthermore, adopting random transmit power selection results in the worst throughput performance among the compared algorithms, until 32 dB.

An important performance metric for all bandit-based algorithms is related to the accumulated regret over the game horizon. Fig. 6 presents the accumulated regret, in terms of throughput for the two versions of BB – PC. It is clear that kl-UCB++ experiences less accumulated regret compared to UCB1, thus revealing that it converges faster to the optimal transmit power level.

VI. CONCLUSIONS

A. Conclusions

Power control is an important technique to guarantee the performance of full-duplex cooperative relay networks. However, the selection of an appropriate power level entails significant coordination overheads while acquiring full channel state information. Targeting to reduce the complexity of this process and provide increased network autonomy, we have
adopted the MAB framework on a stochastic setting with stationary wireless channels and developed relevant power control algorithms. The learning process was based only on ACK/NACK observations, adjusting the relay’s power level in such a way, so as to reduce the impact of loop interference and ensure increased end-to-end throughput. Performance evaluation showed that the proposed algorithms closely follow the optimal power control with full channel state information for different cases of loop interference severity.

B. Future Directions

In this work, we assumed that the channel conditions for different power levels are independent, i.e., no inference for other power levels is made based on the outcome of the trial for a specific power level. In fact they are not, and our method, by assuming this independence, provides the worst case scenario, and converging to the optimal power is slower than what it could be. Part of ongoing work focuses on exploiting the correlation between the outcomes of the trials for different power levels.

Also, the proposed multi-armed bandit-based framework can be applied in a variety of wireless communication areas. The investigation of a multi-relay setup can provide further performance gains to full-duplex transmissions while avoiding excessive coordination overheads through distributed timer-based coordination [22].

At the same time, the consideration of non-stationary channels represents a more practical setting for mobile relay networks and devising efficient learning algorithms should be prioritized.

REFERENCES


