Timer-Based Distributed Channel Access for Control Over Unknown Unreliable Time-Varying Communication Channels

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Timer-Based Distributed Channel Access for Control Over Unknown Unreliable Time-Varying Communication Channels

Tahmoores Farjam, Themistoklis Charalambous, and Henk Wymeersch

Abstract— We consider the resource allocation problem for a system consisting of multiple control subsystems that share an unknown unreliable time-varying wireless channel. We propose a novel method for granting channel access deterministically without requiring information exchange between subsystems, based on local timers for prioritizing channel access with respect to a local cost that takes into account the (initially unknown) channel conditions. We propose a novel setup which has the capability of learning the parameters of imperfect communication links while ensuring distributed collision-free implementation. More specifically, we adopt learning algorithms for estimating the channel quality and, hence, define the local cost as a function of this estimation and control performance. The efficacy of this mechanism is evaluated via simulations.

Index Terms— Wireless networked control systems, distributed channel access, cost of information loss, exploration-exploitation, bandits.

I. INTRODUCTION

Modern control environments, e.g., smart buildings and autonomous vehicles, consist of a large number of spatially distributed smart devices with sensing, actuating and computing capabilities. To perform the control tasks, these devices use a shared wireless medium for communication and comprise systems known as Wireless Networked Control Systems (WNCSs); see, e.g., surveys [1] and [2]. Despite the numerous advantages that result from using a wireless network, for instance, flexibility, installation and maintenance cost reduction, it introduces new challenges which limit its application.

Guaranteeing the control performance of the involved dynamical systems, herein called subsystems, necessitates devising a scheduling mechanism for sharing the wireless network efficiently. This can be achieved by implementing static scheduling mechanisms where the communication sequence is determined off-line, see, e.g., [3], [4]. To capture the dynamic nature of WNCSs and allow for flexibility, dynamic schedulers have been proposed where, at each instance, a scheduling unit grants channel access based on evaluation of some objective function. For instance, [5], [6] use the discrepancy in state estimates as the objective, [7]–[9] base their evaluation on the growth of the standard quadratic

cost (penalizing the error from the deviation from the desired state and the magnitude of the control action), while [10], [11] consider the transmission power consumption in calculations as well.

In many applications, similar to the case considered in this paper, a central scheduling unit is non-existent and thus a distributed resource allocation mechanism is required. In such settings, the definition of the control objective function is restricted to local information only. Moreover, a novel mechanism should be devised by which the subsystems can coordinate for channel access. In this regard, [12] proposed a binary countdown technique which assigns communication channels to the subsystems with the highest state-dependent error. However, the method is prone to collisions in homogeneous systems with limited number of contention slots and requires precise synchronization for explicitly identifying the contention slots. To overcome such shortcomings, [13] introduced the concept of local timers as a means for prioritizing subsystems based on the growth they impose on the quadratic cost and optimally allocate resources in a collision-free manner. Nevertheless, it is designed for perfect communication links and thus fails to account for packet losses which are an inherent feature of the network, due to the uncertain stochastic nature of the wireless medium.

In WNCSs, however, due to the dynamics of the subsystems and the changing environment, the channel conditions vary rapidly under the transmission of a single packet; such systems are fast fading and only statistical Channel State Information (CSI) acquisition is reasonable [14]. Statistical CSI refers to the statistical characterization of the channel, e.g., the type of fading distribution and the average channel gain. Such information is often used for transmission scheduling; the main problem considered in this paper.

In this work, we first provide a distributed solution to the transmission scheduling optimization problem over independent and identically distributed (i.i.d.) packet dropping links by introducing a modified version of the distributed channel access mechanism in [13]. The next major contribution of this work is introduction of a novel method for distributed resource allocation for control over unknown unreliable wireless channels. Unlike the aforementioned works [5]–[9], we assume the realistic scenario where no prior information of the first moment of the statistics of the wireless channels is available. In this case, we cast the channel selection problem as a multi-armed bandit (MAB). To the best of our knowledge, this work presents the first distributed MAB solution for channel access with respect to a control cost in
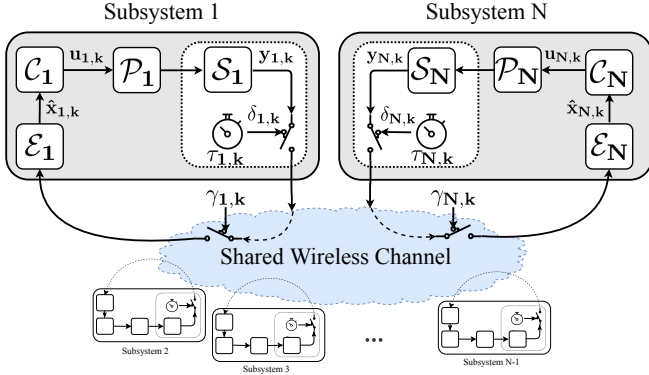
Fig. 1. Example of the WNCS layout where N subsystems compete to access a single wireless channel. $\mathcal{P}_i$ represents the plant of subsystem $i \in \{1, \ldots, N\}$, $\mathcal{S}_i$ its sensor, $\mathcal{E}_i$ its estimator, and $\mathcal{C}_i$ its controller. Note that the timer is embedded in the sensor which is capable of local computations.

WNCSs. We apply the celebrated results obtained for single-player MAB for prioritizing the available channels. Then, by coupling these priorities with the associated control cost of each subsystem in our timer setup, we ensure that channel access is granted in a distributed collision-free manner with respect to the control objective.

The rest of this paper is organized as follows. We first provide the necessary preliminaries and system model in Section II and then briefly describe the distributed channel access mechanism in Section III. In Section IV, we use the multi-player bandit model for learning the unknown channel parameters and propose a novel method for distributed channel access. In Section V, we evaluate how the performance is influenced by employing the proposed method. We draw conclusions and discuss future directions in Section VI.

II. SYSTEM MODEL AND PRELIMINARIES

The setup of the WNCSs under consideration are depicted in Fig. 1, where N decoupled subsystems, which exchange no information between them, share a single wireless channel. Each subsystem $i \in \{1, \ldots, N\}$ consists of a plant ($\mathcal{P}_i$) which receives control inputs from local controller ($\mathcal{C}_i$) and its outputs are measured by the sensor ($\mathcal{S}_i$) which is capable of computations. The resource allocation scheme determines when measurements are transmitted to the corresponding estimator $\mathcal{E}_i$ in a distributed manner.

A. Local Processes

We model the dynamics of each subsystem as linear time-invariant stochastic processes evolving according to

$$
\begin{align}
x_{i,k+1} &= A_i x_{i,k} + B_i u_{i,k} + w_{i,k}, \tag{1a} \\
y_{i,k} &= C_i x_{i,k} + v_{i,k}, \tag{1b}
\end{align}
$$

where $x_{i,k} \in \mathbb{R}^{n_i}$, $y_{i,k} \in \mathbb{R}^{p_i}$ and $u_{i,k} \in \mathbb{R}^{m_i}$ are the local states, measurements and inputs of the actuators of subsystem i at time step k, respectively. Moreover, the stochastic disturbances and measurement noises are assumed to be i.i.d. random sequences described by $w_{i,k} \sim \mathcal{N}(0, W_i)$ and $v_{i,k} \sim \mathcal{N}(0, V_i)$, respectively. Moreover, we assume v and w are mutually independent, i.e., $\mathbb{E}\{wv^T\} = 0$.

B. Imperfect Communication

Let $\mathcal{N}$ and $\mathcal{M}$ denote the index set of subsystems and available channels, respectively, with $|\mathcal{N}| = N$ and $|\mathcal{M}| = M$, where the notation $|\cdot|$ in the case of sets stands for the cardinality of the set. While the communication link between the local controller and actuators is assumed to be perfect, the sensors transmit measurements over lossy wireless links to be received by their corresponding estimators. We assume a time-slotted medium access protocol is implemented and let the decision variable $\delta_{i,j,k} \in \{0, 1\}$ represent whether subsystem i transmits on channel j at time step k or not as follows

$$
\delta_{i,j,k} = \begin{cases} 1, & y_{i,k} \text{ is transmitted on channel } j, \\ 0, & \text{otherwise.} \end{cases}
$$

Since wireless channels are unreliable, $\delta_{i,j,k} = 1$ does not guarantee successful reception of the data packet. Hence, we define an additional binary variable $\gamma_{i,j,k}$ to denote the confirmation of successful delivery as

$$
\gamma_{i,j,k} = \begin{cases} 1, & y_{i,k} \text{ is successfully received over channel } j, \\ 0, & \text{otherwise,} \end{cases}
$$

which represents the acknowledgment signal. Typically, the problem of scheduling arises when the shared communication resources are limited, as is the case for the considered WNCS (i.e., $M < N$). To avoid collisions, at a given time, a channel can only be used by one subsystem

$$
\sum_{i \in \mathcal{N}} \delta_{i,j,k} \leq 1, \qquad \forall j, \forall k. \tag{2}
$$

Moreover, each subsystem can only use one channel at a given time

$$
\sum_{j \in \mathcal{M}} \delta_{i,j,k} \leq 1, \qquad \forall i, \forall k. \tag{3}
$$

We model packet delivery over each channel as i.i.d. Bernoulli random sequences. Therefore, the probability of successful transmission satisfies a Bernoulli distribution with mean $q_{i,j} \in [0, 1]$. Using the introduced decision variables, the expectation of successful packet delivery is given by

$$
\mathbb{P}(\gamma_{i,j,k} = 1 | \delta_{i,j,k} = 1) = q_{i,j}. \tag{4}
$$

C. Controller and Estimator

In this work, the standard quadratic cost over the infinite horizon is chosen as the performance measure, which we intend to minimize. This cost is defined as

$$
J_0 = \mathbb{E}\left\{ \lim_{\kappa \to \infty} \frac{1}{\kappa} \sum_{k=0}^{\kappa-1} \sum_{i=1}^{N} \left(x_{i,k}^T Q_i x_{i,k} + u_{i,k}^T R_i u_{i,k} \right) \right\}, \tag{5}
$$

where Q_i and R_i are positive definite weighing matrices of appropriate dimensions. The control inputs for minimizing this cost are computed by

$$
u_{i,k} = L_i \hat{x}_{k|k}, \tag{6}
$$

where L_i is the stabilizing feedback matrix given by

$$
L_i = -(B_i^T \Pi_i B_i + R_i)^{-1} B_i^T \Pi_i A_i, \tag{7}
$$

where Π_i is the solution of discrete-time algebraic Riccati equation. Furthermore, using Kalman filter as the local estimator, the *a posteriori* state estimates, denoted by $\hat{x}_{i,k|k}$, can be calculated recursively by [15]

$$\hat{x}_{i,k|k-1} = (A_i + B_i L_i)\hat{x}_{i,k-1|k-1}, \tag{8a}$$

$$P_{i,k|k-1} = A_i P_{i,k-1|k-1} A_i^T + W_i, \tag{8b}$$

$$K_{i,k} = P_{i,k|k-1} C_i^T \left(C_i P_{i,k|k-1} C_i^T + V_i \right)^{-1}, \tag{8c}$$

$$\hat{x}_{i,k|k} = \hat{x}_{i,k|k-1} + (\textstyle\sum_{j=1}^{M} \gamma_{i,j,k}) K_{i,k}(y_{i,k} - C_i \hat{x}_{i,k|k-1}), \tag{8d}$$

$$P_{i,k|k} = (I - (\textstyle\sum_{j=1}^{M} \gamma_{i,j,k}) K_{i,k} C_i) P_{i,k|k-1}, \tag{8e}$$

where the covariance matrices and Kalman gain are denoted by P and K, respectively. Initial conditions $P_{i,0}$ and $\hat{x}_{i,0}$ are often chosen arbitrarily, or, based on prior knowledge from available information by the respective process [16]. Note that for each subsystem $i \in \mathcal{N}$, $\sum_{j=1}^{M} \gamma_{i,j,k} \leq 1$, since i will transmit to one of the available channels at a given time slot.

D. Timer-based Mechanism

The timer-based mechanism introduced in [13], denoted by TBCoIL, provides a distributed solution for the channel allocation problem in settings similar to our WNCS structure. This mechanism is based on the idea of assigning local timers to every subsystem. At each time step k, the timers are set to

$$\tau_{i,k} = \frac{\lambda}{m_{i,k}}, \tag{9}$$

where λ is a constant shared among all subsystems, which can be fine-tuned for establishing the desired contention period. Furthermore, $m_{i,k}$ is a measure of error which can be chosen arbitrarily. By choosing a cost that can be calculated according to the locally available information only as the measure $m_{i,k}$, all subsystems can compute (9) independently. Since $\tau_{i,k}$ is inversely proportional to this cost, timers can be viewed as a means for prioritizing channel access. At the beginning of each transmission slot, the timers are set and started. The subsystem that possesses the smallest timer, i.e., the largest cost, has the highest priority for claiming the channel. As the first timer expires, the corresponding subsystem sends a short-duration flag packet on the network which informs the remaining contestants to stop their timers and back off. Then, data packet is transmitted during this slot without any collision. As the next transmission slot begins, the timers are reset to new values and the same procedure is repeated.

III. DISTRIBUTED CHANNEL ACCESS MECHANISM

Here, we first modify TBCoIL to extend its application to the case where multiple fading channels are present. This can be achieved by assigning multiple timers to each subsystem. More specifically, each subsystem possesses M independent timers which correspond to M available channels. Similar to the original approach, (9) is used for calculating $\tau_{i,j,k}$ where the additional index $j \in \mathcal{M}$ corresponds to the considered channel. Moreover, the local cost, denoted by $m_{i,j,k}$, is calculated individually for each channel.

In this framework, at any given time step k, the first claimed channel j^* and the corresponding subsystem i^* are determined by the first timer that expires (reaches zero), i.e., $\{i^*, j^*\} = \arg\min_{i,j}\{\tau_{i,j,k}\}$. This subsystem transmits a short duration flag packet on channel j^* immediately, thus informing all other subsystems in the network to stop their timers for this channel and back off. Simultaneously, i^* stops its remaining timers and thus withdraws from competition for the remaining channels and starts to transmit on channel j^* without collision. Meanwhile, the rest of subsystems compete for available resources until all M channels have been allocated for. Similar to the original method, as the time slot ends, all timers are reset and the whole procedure is repeated.

Remark 1: Although, for ease of explanation, we assume each subsystem possesses multiple timers, a single clock suffices for implementation. The value of an imaginary timer assigned to a specific channel can equivalently be represented by a checkpoint on the elapsed time of the clock from the beginning of that time slot. Consequently, as the clock hits the first checkpoint, it can be interpreted as the smallest timer reaching zero. Therefore, the corresponding channel is claimed and all the remaining checkpoints are removed. Furthermore, if a flag packet is received, the checkpoint of that channel is neglected which is equivalent to backing off and collision avoidance.

A. Timer setup

The main challenge in implementation of TBCoIL is quantification of the local cost. In order to allow for distributed implementation, this cost should be defined in a manner that it could be determined by locally available information only. In addition, our main goal is to minimize (5). Hence, the local cost should be defined such that by implementing the resulting channels access scheme, our goal is attained.

Optimal resource allocation results in minimization of (5). Let $\mathcal{S}_k \subseteq \mathcal{N}$ denote the set of subsystems that transmit their measurement at k and define $\bar{\mathcal{S}}_k \triangleq \mathcal{N} \backslash \mathcal{S}_k$. Furthermore, we define $E_{i,k}^0$ as a cost incurred by subsystem i in case it does not receive measurement updates at k; similarly, $E_{i,k}^1$ is the cost when this subsystem receives the observations. Hence, over perfect communication links, the total cost at step k can be written as [8]

$$J_k = \sum_{i \in \bar{\mathcal{S}}_k} E_{i,k}^0 + \sum_{i \in \mathcal{S}_k} E_{i,k}^1. \tag{10}$$

This cost can be defined for communication over unreliable channels in a similar manner [8]. Let $\bar{j}_i : \mathcal{S}_k \to \{1, \dots, M\}$ denote the index j with $\delta_{i,j,k} = 1$ for $i \in \mathcal{S}_k$, i.e., the index of the channel allocated to subsystem i for transmission. The cost in this setup is defined as [8]

$$\mathbb{E}\{J_k | \mathcal{S}_k\} = \sum_{i \in \bar{\mathcal{S}}_k} E_{i,k}^0 + \sum_{i \in \mathcal{S}_k} \left(E_{i,k}^0 (1 - q_{i,\bar{j}_i}) + E_{i,k}^1 q_{i,\bar{j}_i} \right)$$
$$= \sum_{i \in \mathcal{N}} E_{i,k}^0 + \sum_{i \in \mathcal{S}_k} \left(E_{i,k}^1 - E_{i,k}^0 \right) q_{i,\bar{j}_i}$$

$$= \sum_{i \in \mathcal{N}} E_{i,k}^0 - \sum_{i \in \mathcal{S}_k} \text{CoIL}_{i,k} q_{i,\bar{j}_i}, \quad (11)$$

where $\text{CoIL} \triangleq E_{i,k}^0 - E_{i,k}^1$ denotes the *cost of information loss*. This cost can be construed as the increase in the quadratic cost imposed by a subsystem, in case it does not receive measurement updates. Minimizing (11) is equivalent to maximizing the last term which can be interpreted as the optimal resource allocation problem at k. This problem can be formulated as

$$\max_{\delta_{i,j,k} \in \{0,1\}} \sum_{i=1}^{N} \sum_{j=1}^{M} \text{CoIL}_{i,k} q_{i,j} \delta_{i,j,k}, \quad (12)$$

constrained by (2) and (3). We aim at solving this problem in a distributed manner. As aforementioned, if local information is sufficient for determining the cost, implementing the timer-based mechanism ensures that channel access is granted to subsystems with the highest cost in a distributed fashion.

It has been proven in [8] that by considering the quadratic cost (5), for decoupled dynamics, we can derive

$$\text{CoIL}_{i,k} = \text{tr}\left(\Gamma_i \left(P_{i,k|k-1} - P_{i,k|k}\right)\right), \quad (13)$$

where Γ_i is a weighting matrix and $P_{i,k|k-1}$ and $P_{i,k|k}$ are the *a priori* and *a posteriori* error covariance matrices as defined in (8b) and (8e), respectively. Since all the needed parameters for determining this cost are independent of the measurements and all needed is the initial condition, the updates of the error covariance matrices can be computed locally and away from the sensor that takes the measurements. As a result, the product in (12) can be used as the local measure for timer setup under the assumption that channel parameters are known. Hence, by setting the timers to

$$\tau_{i,j,k} = \frac{\lambda}{\text{CoIL}_{i,k} q_{i,j}}, \quad (14)$$

the first M timers which expire determine the transmitting subsystems and the corresponding claimed channels. As a result, utilizing this setup provides a distributed solution for (12) and guarantees collision-free channel access, provided that there are no collisions between the flag signals broadcasted by the subsystems.

IV. CHANNEL ACCESS OVER UNKNOWN CHANNEL STATISTICS

A. Problem Statement

Optimal resource allocation requires knowledge of the exact values of $q_{i,j}$. In slow fading channels, this parameter can be approximated with estimation processes, e.g., instantaneous CSI acquisition. However, due to the dynamic nature of the considered subsystems and the changing environment, the coherence time of the channel is relatively small and thus fast fading occurs. In such settings, learning methods can be applied as an alternative for gaining knowledge of the underlying channel parameters.

Despite the abundance of existing learning algorithms, adopting a suitable one is challenging for two setup-related reasons: *(i)* our WNCS structure allows no information exchange among subsystems and thus the learning should be accomplish in a distributed way; *(ii)* since the main objective is minimizing the overall cost, the adopted algorithm should be compatible with the proposed timer-based mechanism. More specifically, channel statistics cannot be learned separately without taking into consideration CoIL. This is due to the fact that our main objective is to enable the subsystems to optimally perform their control tasks and ignoring CoIL would result in deterioration in control performance. We aim at devising a novel distributed method which takes into consideration the trade-off between learning the channel statistics and control performance to address the aforementioned challenges.

B. A MAB Approach

MAB problem refers to optimal sequential allocation in unknown random environments. Due to the various advantages they offer, there has been a recent surge in extending the application of MAB methods to the field of wireless communication in areas, such as relay selection [17], opportunistic spectrum access [18], cognitive radio systems [19] and energy-efficient small cells planning in 5G networks [20]. In classic single-player stochastic MAB, a player has access to multiple, say M, independent arms. At each round, this player can pull an arm $j \in \{1, \ldots, M\}$ which yields a random reward from an unknown probability distribution specific to that arm. Since the player has no prior knowledge of the reward distribution, he might play an inferior arm in terms of reward. We define *regret* as the difference between the reward achieved when the best arm is pulled and the player's choice. Let $r_{j,k}$ and I_k denote the instantaneous reward obtained from arm j and the selected arm at round k, respectively. Then, we define the (expected) regret up to round κ as

$$R_\kappa = \max_{j \in \{1, \ldots, M\}} \mathbb{E}\left[\sum_{k=1}^{\kappa} r_{j,k}\right] - \mathbb{E}\left[\sum_{k=1}^{\kappa} r_{\text{I}_k}\right]. \quad (15)$$

The objective is to find a *policy* for selecting the arms in a way that (15) is minimized over the game horizon. The performance of a policy relies on how it addresses *exploration/exploitation* dilemma: searching for a balance between exploring all arms to learn their reward distribution while playing the best arm more often to gain more reward.

The channel selection procedure of a subsystem can conveniently be cast as a single-player MAB. In this scenario, channels represent arms and pulling an arm corresponds to packet transmission over the selected channel. We assume a binary rewarding scheme ($r_{j,k} \in \{0, 1\}$), where in case of successful reception of the data packet, a unit reward is obtained over that channel ($r_{\text{I}_k} = 1$), whereas failed transmissions correspond to pulling an arm which offers no reward ($r_{\text{I}_k} = 0$). The channels are independent and packet dropouts are i.i.d. random sequences, or in other words, the rewards are i.i.d. random. The mean of the Bernoulli distribution of rewards over each channel correspond to the probability of successful transmission (4). Therefore, by adopting suitable policies, after the initial exploration phase,

the channel with the best quality is exploited for maximizing the success rate or, equivalently, the reward.

A class of solutions to this problem are known as *index policies*, which assign an index to each arm and the one with the largest index is played. One of the main categories of methods that belong to this class are based on *upper confidence bound* (UCB). These policies estimate an upper bound of the mean reward $\bar{q}_{j,k}$ of each arm j at some fixed confidence level and determine the indices according to the estimated bounds. One of the celebrated results based on this idea is UCB1 policy introduced in [21]. At each round k, UCB1 first computes the upper confidence bound of the mean reward, $\bar{q}_{j,k}$, as follows

$$\bar{q}_{j,k} = \hat{q}_{j,k} + \sqrt{\frac{2 \ln z_k}{z_{j,k}}}, \qquad (16)$$

where z_k is the total number of plays, $z_{j,k}$ denotes the number of plays of arm j and $\hat{q}_{j,k}$ is the average reward obtained from playing this arm up to k, i.e.,

$$\hat{q}_{j,k} = \frac{\sum_{s=1}^{k} r_{j,s} 1_{I_s=j}}{z_{j,k}}, \qquad (17)$$

where $1_{I_k=j}$ is 1 when the index at time k, I_k, is assigned to arm j. Then, once the upper confidence bound of the mean reward is obtained for all arms, UCB1 assigns index I_k to the arm with the maximum upper confidence bound, i.e.,

$$I_k = \arg\max_{j \in \{1,\ldots,M\}} \bar{q}_{j,k}. \qquad (18)$$

In addition to computational efficiency, under some mild assumptions, this policy is order optimal over time [21].

The problem of distributed channel access in standard wireless networks concerns maximizing the number of successful transmissions, which can be cast as a multi-player MAB. To define regret, we first determine the maximum obtainable reward. Using the aforementioned binary rewarding scheme, maximizing the reward at every time step k corresponds to optimal resource allocation, which is

$$\max_{\delta_{i,j,k} \in \{0,1\}} \sum_{i=1}^{N} \sum_{j=1}^{M} q_{i,j} \delta_{i,j,k}, \qquad (19)$$

subject to constraints (2) and (3). Since the reward distribution over each wireless link is assumed to be time-invariant, the optimal decision variables are likewise time-invariant. Consequently, subscript k is dropped and we denote the solution by $\delta_{i,j}^{q*}$. As a result, regret is defined as

$$R_\kappa = \kappa \sum_{i=1}^{N} \sum_{j=1}^{M} q_{i,j} \delta_{i,j}^{q*} - \sum_{k=1}^{\kappa} \sum_{i \in \mathcal{S}_k} q_{i,\bar{j}_i}, \qquad (20)$$

which can be minimized by adopting distributed policies, such as UCB1.

C. Distributed Channel Access Algorithm

In this section, we describe our proposed distributed channel access mechanism. We first cast our problem as a multi-player MAB. Then, we propose a novel indexing policy for addressing the exploration/exploitation dilemma in a collision-free distributed manner. More specifically, we use time-varying weights, which reflect the control performance, in index calculations. Furthermore, by implementing the timers for coordination, we ensure that the players can access the channels without collision. The detailed description of the algorithm follows.

Since our goal is to minimize (5), with a slight abuse of notation, we define the *cost regret* up to time step κ as

$$R_{cost,\kappa} = \sum_{k=1}^{\kappa} \mathbb{E}\{J_k\} - \sum_{k=1}^{\kappa} \mathbb{E}\{J_k^*\}, \qquad (21)$$

where $\mathbb{E}\{J_k^*\}$ is the minimum of (11), which can be achieved by optimal resource allocation, i.e., (12). Although the cost regret fundamentally differs from the standard regret defined in (20), we propose a new method for exploiting the well-established results for minimizing the latter in our favor.

While now the regret function has changed, we still apply the aforementioned binary rewarding scheme in, e.g., (16), assuming that at any given time k, in case subsystem i transmits on channel j ($\delta_{i,j,k} = 1$) and the transmission is successful ($\gamma_{i,j,k} = 1$), a unit reward is obtained. Nevertheless, the indices are assigned with respect to the weighted expected reward, i.e.,

$$I_{i,k} = \begin{cases} \arg\max_{j \in \{1,\ldots,M\}} \delta_{i,j,k}^*, & \text{if } \exists j : \delta_{i,j,k}^* \neq 0, \\ \emptyset, & \text{otherwise,} \end{cases}$$

where $\delta_{i,j,k}^*$ is obtained by optimization

$$\boldsymbol{\delta}_k^* \triangleq \begin{bmatrix} \delta_{i_1,1,k}^* & \cdots & \delta_{i_M,M,k}^* \end{bmatrix}^T$$
$$= \arg\max_{\delta_{i,j,k} \in \{0,1\}} \sum_{i=1}^{N} \sum_{j=1}^{M} \text{CoIL}_{i,k} \bar{q}_{i,j,k} \delta_{i,j,k}, \qquad (22)$$

subject to the channel constraints (2) and (3). This ensures correct estimation of the success probability of the channel, while at the same time the slot is allotted (in a distributed fashion) to the subsystem with the highest cost in order to minimize the cost regret given in (21). This policy can be enforced by implementing the timer-based mechanism for solving (22). By using the weighted expected reward as the local measure, (9) can be rewritten as

$$\tau_{i,j,k} = \frac{\lambda}{\text{CoIL}_{i,k}\, \bar{q}_{i,j,k}}. \qquad (23)$$

As a result, assuming a negligible flag packet duration, this mechanism provides collision-free channel access.

In case of standard wireless networks where regret is defined as (20), in addition to UCB1, several distributed algorithms have been proposed for improving performance, e.g., *rhoRand*, *RandTopM* and *MCTopM* [19]. Nevertheless, all the solutions with respect to (20) only maximize the number of successful transmission disregarding the dynamics and

control performance, which are the critical characteristics of WNCSs. Despite the ineffectiveness of their standalone implementation in our setup, manipulating their outcome in our policy by the time-varying weights, i.e., CoIL, results in significant improvements as shown by the numerical results in Section V. Algorithm 1 illustrates the detailed distributed implementation of our policy at each subsystem when UCB1 is used as the core policy.

Remark 2: UCB1 algorithm requires each arm to be pulled at least once when initiating. For this reason, at the beginning we adopt a pre-specified schedule in which all N subsystems transmit to all M channels. This procedures requires N steps. Afterwards, the generated set of observations and cumulative rewards, denoted by $\mathcal{Z}_{i,1} \triangleq \{z_{i,j,1} | \forall j \in \mathcal{M}\}$ and $\mathcal{R}_i \triangleq \{R_{i,j} | \forall j \in \mathcal{M}\}$, respectively, are used for determining channel access according to (22).

Algorithm 1: Timer-based channel access mechanism for subsystem i

---

Input: number of channels M, constant value for timer setup λ, the set of observation history $\mathcal{Z}_{i,1}$ and cumulative rewards $\mathcal{R}_i$.

1 for $k = 1, 2, \ldots$ **do**
2 $z_{i,k} = \sum_{j=1}^{M} z_{i,j,k}$ and $\hat{q}_{i,j,k} = \frac{R_{i,j}}{z_{i,j,k}}$
3   calculate $\bar{q}_{i,j,k}$ (16) and $\text{CoIL}_{i,k}$ (13)
4   start timers $\tau_{i,j,k} = \frac{\lambda}{\text{CoIL}_{i,k} \, \bar{q}_{i,j,k}}$ (24)
5   initiate set of dummy indices $\mathcal{F}_i = \{1, \ldots, M\}$
6   **while** $|\mathcal{F}_i| > 0$ **do**
7   **for** $j \leftarrow 1$ **to** M **do**
8   **if** $\tau_{i,j,k} \neq 0$ *and timer is running* **then**
9   listen for signals
10   **if** *signal is received in channel j* **then**
11   freeze $\tau_{i,j,k}$ and set $f_{i,j} = \emptyset$
12   **end**
13   **else if** $\tau_{i,j,k} = 0$ **then**
14   send flag on channel j
15   set $\mathcal{F}_i = \emptyset$ and $\text{I}_{i,k} = j$
16   freeze all running timers
17   **end**
18   **end**
19   **end**
20 $\mathcal{Z}_{i,k+1} = \mathcal{Z}_{i,k}$
21   **if** $\text{I}_{i,k} \neq \emptyset$ **then**
22   transmit on channel $\text{I}_{i,k}$
23 $z_{i,\text{I}_{i,k},k+1} = z_{i,\text{I}_{i,k},k+1} + 1$
24   **if** $\gamma_{i,\text{I}_{i,k},k} = 1$ **then** $R_{i,\text{I}_{i,k}} = R_{i,\text{I}_{i,k}} + 1$
25   **end**
26 end

Remark 3: Although we adopted UCB1 in our proposed setup due to its low computational complexity, our algorithm provides the flexibility of using any indexing policy, which has the capability of distributed implementation, as the measure of channel quality.

V. NUMERICAL RESULTS

The following results have been obtained for WNCSs consisting of two classes of homogeneous dynamical systems. The unstable subsystems form class I, while class II consists of stable subsystems. We consider the scenario where

$$A_{\text{I}} = \begin{bmatrix} 1.2 & 0 \\ 0 & 1.1 \end{bmatrix} A_{\text{II}} = \begin{bmatrix} 0.9 & 0.1 \\ 0 & 0.9 \end{bmatrix}, B = C = I_{2 \times 2},$$

where I denotes the identity matrix. The state estimates and feedback control law are determined as discussed and we intend to minimize the quadratic cost defined in (5) with $Q = I_{2 \times 2}$ and $R = 0.01 I_{2 \times 2}$.

Example 1: Here, we demonstrate how the resources are allocated in a small WNCS where three subsystems compete for two available communication channels. Subsystem 1 is chosen from class II while the rest are chosen from class I, and hence they are unstable. The quality of each wireless link for each subsystem is assumed to be according to Table I.

TABLE I
PROBABILITY OF SUCCESSFUL TRANSMISSION OVER EACH CHANNEL
FOR SUBSYSTEMS OF EXAMPLE 1

Subsystem	1	2	3
Channel 1	0.4	0.8	0.6
Channel 2	0.7	0.5	0.9

Fig. 2 depicts how the resources are allocated when different costs are used as the local measure in (9). For the first case, we use a slightly modified version of TBCoIL and set the timers to

$$\tau_{i,j,k} = \frac{\lambda}{\text{CoIL}_{i,k} \, q_0}, \tag{24}$$

where $q_0 = 0.5$. This corresponds to assuming a 50% chance of successful transmission over every wireless link. As a result, when a timer reaches zero, the corresponding sensor selects one of the available channels randomly for transmission. As it can be seen, using this setup, the channels are occupied by unstable subsystems more often. However, the channels are selected randomly with no regards to the real probability of successful transmission.

For the next case, the index assigned by UCB1, denoted by $\bar{q}$, is chosen as the local measure. In order to avoid collisions, we assume all subsystems can communicate with a central scheduler. Therefore, if the timers on the same channel expire simultaneously for multiple subsystems, the scheduler assigns the channel to one of them randomly. After the initial exploration phase, Subsystems 2 and 3 exploit Channels 1 and 2, respectively. Since only the channel quality determines the priorities, resources are dedicated to subsystems which have the highest probability of successful transmission according to the past observations. Therefore, mainly the two aforementioned subsystems claim the channels. Implementing CoIL$\bar{q}$ in (9), however, leads to a combination of the outcomes of the previous cases. The unstable subsystems claim the channels more often, while they tend to occupy the channel which offers the highest probability of success.
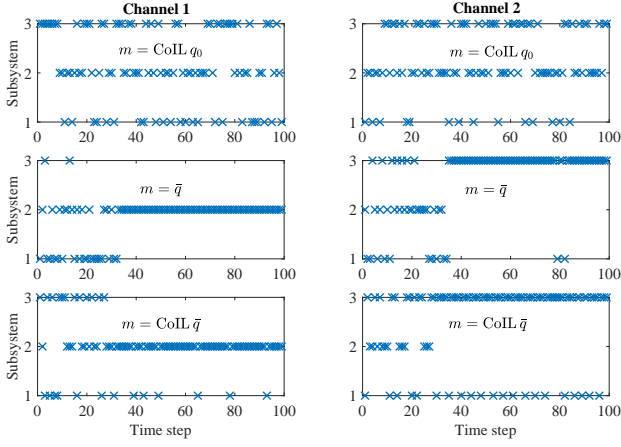
Fig. 2. Resource allocation for the case of three subsystems competing over two available channels using timers set by various measures.

The solution to the optimal resource allocation problem (12) can be interpreted as pairing the subsystems with the largest CoIL with channels with highest probability of success. Hence, intuitively, it is expected that Subsystems 3 and 2 transmit more often on Channels 3 and 2, respectively, since they offer a higher chance of successful packet delivery. Furthermore, due to its stability in addition to lower success rate of wireless links, Subsystem 1 is expected to occupy the channels less frequently. The results obtained for the last case, where the timers are set to (23), follow this intuition closely.

The performance of the discussed setups in terms of average regret over a large horizon is depicted in Fig. 3. Here, we consider two additional setups where the channel parameters are known *a priori* and a central scheduler allocates the resources for maximizing the reward. These setups lead to the optimal solution of (19) and (12), where the local measure is set to q and CoILq, respectively. As it can be seen, the regret while using the setup with UCB1 indices as the local measure converges to zero. Moreover, allocating the resources without considering the channel statistics, using (24), leads to the largest regret. However, considering the cost regret, this setup outperforms the cases where control performance is neglected. Moreover, the average cost regret of our proposed policy converges to zero fast indicating satisfactory performance.

Remark 4: The unstable subsystems require communication resources more frequently and since the indices are assigned with respect to CoIL, they are chosen for transmission *infinitely often*[1] and hence eventually the channel statistics are learned accurately. In contrast, stable subsystems rarely transmit and thus they might learn the statistics with less accuracy. Subsequently, the unknown parameters which are essential for guaranteeing control performance are learned accurately, which implies that over time, the resources are allocated similarly to the optimal case where the channel

---

[1]Let $\{A_n\}_{n=1}^{\infty}$ be an infinite sequence of events. We say that events in the sequence occur "infinitely often" if A_i holds true for an infinite number of indices $i \in \{1, 2, \ldots\}$. See also Borel-Cantelli lemma.

statistics are known. Although this approach is reminiscent of those in [22], [23], in our case, the rewards are weighted by a function of the corresponding covariance matrix, namely CoIL, and due to the time-varying nature of CoIL providing bounds on convergence rate is still a challenging open problem.
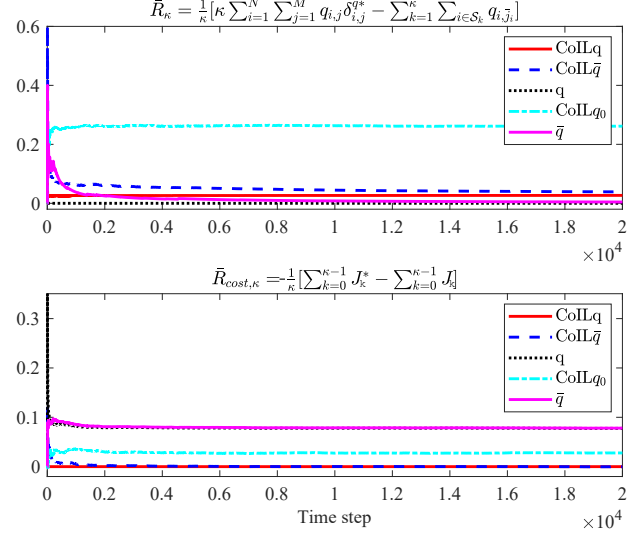

Fig. 3. Average regret and cost regret of three subsystems competing for two available wireless channels. Each line corresponds to a different measure used in the timer setup.

Example 2: Here, we use the modified version of TB-CoIL (24) as the benchmark for comparing different scenarios. We demonstrate the performance of our proposed method in real applications where the channel parameters are unknown by implementing various indexing policies. More specifically, we use four different UCB algorithms, namely UCB1 and UCB-Tuned [21], kl-UCB [24] and UCB-V [25].

Here, we consider WNCSs where half of the subsystem are from class I and the rest are stable and only $1/4$ of them can transmit at each time step. Fig. 4 shows the percentage of cost reduction achievable by different setups compared to the benchmark (24). The depicted results are the average over 50 simulations and the rewards in UCB algorithms are offered similar to Example 1. Furthermore, we model the unreliability of each communication link as random dropouts satisfying a Bernoulli distribution with a unique random mean $q_{i,j} \in [0.4, 0.8]$ for each user.

As expected, the best cost reduction is achieved when the mean of the reward distribution, i.e., the expectation of successful transmission, is known. This setup results in 9.75% up to 18.76% lower cost compared to applying (24). These results are closely followed when UCB algorithms are adopted for prioritizing channel access. Although these algorithms have specific features and growth rate of regret in original MAB problems, their performance is not considerably different in the timers framework. However, all of them lead to significant cost reduction compared to the case where no learning is applied.
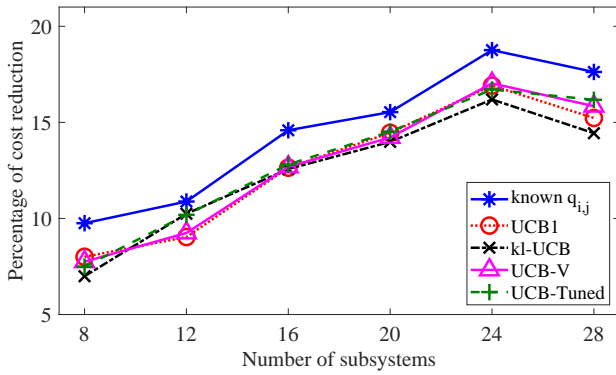
Fig. 4. Reduction in the standard quadratic cost achieved by applying various core policies in timer-base mechanism compared to (24) when $M = N/4$.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

A. Conclusions

In this paper, we presented a novel solution for distributed resource allocation in WNCSs with imperfect communication links. We adopted the concept of local timers for prioritizing channel access and proposed a novel method for improving performance when channel parameters are unknown. We cast the optimal resource allocation problem as a MAB and proposed a novel policy for granting channel access. This policy utilizes well-known indexing policies for estimating the success probability of channels and weighs them by a time-varying control measure, namely CoIL. These weighted values are then used for setting the timers in the timer-based channel access mechanism. As a result, channel parameters are learned while desired control performance is ensured. The simulations showed that the best performance with the timer-based mechanism is achieved when the channel parameters are known *a priori*. When the parameters are unknown, however, implementing our proposed policy leads to significant improvements when compared to policies in which the channel conditions are ignored.

B. Future Directions

Interesting future research directions include investigating the case where the duration of flag packet is not negligible. Therefore, the possibility of collisions in the flag signals can lead to collision during data transmission. Furthermore, this work can be extended to the case where temporal correlation of the variation in channel states are considered.

REFERENCES

[1] R. A. Gupta and M. Y. Chow, "Networked Control System: Overview and Research Trends," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 7, pp. 2527–2535, July 2010.

[2] P. Park, S. C. Ergen, C. Fischione, C. Lu, and K. H. Johansson, "Wireless Network Design for Control Systems: A Survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 978–1013, Secondquarter 2018.

[3] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, Jan 2007.

[4] W. Zhang, M. S. Branicky, and S. M. Phillips, "Stability of networked control systems," *IEEE Control Systems*, vol. 21, no. 1, pp. 84–99, Feb 2001.

[5] D. Han, J. Wu, H. Zhang, and L. Shi, "Optimal sensor scheduling for multiple linear dynamical systems," *Automatica*, vol. 75, pp. 260 – 270, 2017.

[6] C. Li and N. Elia, "Stochastic sensor scheduling via distributed convex optimization," *Automatica*, vol. 58, pp. 173 – 182, 2015.

[7] A. Molin, C. Ramesh, H. Esen, and K. H. Johansson, "Innovations-based priority assignment for control over CAN-like networks," in *IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 4163–4169.

[8] T. Charalambous, A. Ozcelikkale, M. Zanon, P. Falcone, and H. Wymeersch, "On the Resource Allocation Problem in Wireless Networked Control Systems," in *IEEE Conference on Decision and Control (CDC)*, Dec 2017, pp. 4147–4154.

[9] M. Zanon, T. Charalambous, H. Wymeersch, and P. Falcone, "Optimal Scheduling of Downlink Communication for a Multi-Agent System With a Central Observation Post," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 37–42, Jan 2018.

[10] K. Gatsis, M. Pajic, A. Ribeiro, and G. J. Pappas, "Opportunistic Control Over Shared Wireless Channels," *IEEE Transactions on Automatic Control*, vol. 60, no. 12, pp. 3140–3155, Dec. 2015.

[11] A. S. Leong, S. Dey, and D. E. Quevedo, "Sensor scheduling in variance based event triggered estimation with packet drops," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1880–1895, April 2017.

[12] M. H. Mamduhi, A. Molin, D. Tolic, and S. Hirche, "Error-dependent data scheduling in resource-aware multi-loop networked control systems," *Automatica*, vol. 81, pp. 209–216, 2017.

[13] T. Farjam, T. Charalambous, and H. Wymeersch, "A Timer-Based Distributed Channel Access Mechanism in Networked Control Systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 5, pp. 652–656, May 2018.

[14] X. Liu and A. Goldsmith, "Wireless network design for distributed control," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 3. IEEE, 2004, pp. 2823–2829.

[15] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, Sept 2004.

[16] E. Suzdaleva, "Initial Conditions for Kalman Filtering: Prior Knowledge Specification," in *Proceedings of the WSEAS International Conference on Systems Theory and Scientific Computation*. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2007, pp. 45–49.

[17] S. Maghsudi and S. Stanczak, "Transmission mode selection for network-assisted device to device communication: A levy-bandit approach," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 7009–7013.

[18] H. Liu, K. Liu, and Q. Zhao, "Learning in a changing world: Restless multiarmed bandit with unknown dynamics," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1902–1916, March 2013.

[19] L. Besson and E. Kaufmann, "Multi-player bandits revisited," in *Proceedings of Algorithmic Learning Theory*, vol. 83, Apr 2018, pp. 56–92.

[20] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5g small cells," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 64–73, June 2016.

[21] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2, pp. 235–256, May 2002.

[22] A. Carpentier, A. Lazaric, M. Ghavamzadeh, R. Munos, and P. Auer, "Upper-confidence-bound algorithms for active learning in multi-armed bandits," in *International Conference on Algorithmic Learning Theory*. Springer, 2011, pp. 189–203.

[23] A. Carpentier and R. Munos, "Finite time analysis of stratified sampling for monte carlo," in *Advances in Neural Information Processing Systems*, 2011, pp. 1278–1286.

[24] O. Cappé, A. Garivier, O.-A. Maillard, R. Munos, and G. Stoltz, "Kullback-Leibler upper confidence bounds for optimal sequential allocation," *Ann. Statist.*, vol. 41, no. 3, pp. 1516–1541, 06 2013.

[25] J. Y. Audibert, R. Munos, and C. Szepesvári, "Exploration–exploitation tradeoff using variance estimates in multi-armed bandits," *Theoretical Computer Science*, vol. 410, no. 19, pp. 1876–1902, 2009.