Learning to Speak on Behalf of a Group

Medium Access Control for Sending a Shared Message

Haque, Shaan ul; Chandak, Siddharth; Chiariotti, Federico; Günduz, Deniz; Popovski, Petar

Published in:
IEEE Communications Letters

DOI (link to publication from Publisher):
10.1109/LCOMM.2022.3181733

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

Learning to Speak on Behalf of a Group: Medium Access Control for Sending a Shared Message

Shaan ul Haque, Siddharth Chandak, Federico Chiariotti, Deniz Gündüz, and Petar Popovski

***Abstract*—The rapid development of Internet of Things (IoT) technologies has not only enabled new applications, but also presented new challenges for reliable communication with limited resources. In this work, we define a novel problem that can arise in these scenarios, in which a set of sensors need to communicate a joint observation. This observation is shared by a random subset of the nodes, which need to propagate it to the rest of the network, but coordination is complex: as signaling constraints require the use of random access schemes over shared channels, sensors need to implicitly coordinate, so that at least one transmission gets through without collisions. Unlike the majority of existing medium access schemes, the goal is to make sure that the shared message gets through, regardless of the sender. We analyze this coordination problem theoretically and provide low-complexity solutions. While a clustering-based approach is near-optimal if the sensors have prior knowledge, we provide a distributed multi-armed bandit (MAB) solution for the more general case and validate it by simulation.**

***Index Terms*—Distributed coordination, multi-armed bandit, Thompson sampling, random access**

I. Introduction

Over the past few years, the rise of the Internet of Things (IoT) [1] has opened new possibilities in the manufacturing, energy, and health sectors. The promise of 5G and beyond networks is to support massive numbers of sensors and machine-type devices, along with sporadic low-latency communications, without affecting human communication traffic [2]. However, there are still many open problems in coordinating medium access for sporadically active sensors [3], particularly in remote deployments with very limited resources. As random access schemes like slotted ALOHA are entirely distributed, and only require limited overhead for slot synchronization, they often represent the best choice for these scenarios [4], but the risk of collisions is a significant drawback, particularly for scenarios with a large number of nodes [5].

An interesting scenario in this context is the transmission of an observation or a decision that is shared by the devices in a given *active set*. For example, this could be the position of a target object, or an abnormal value of a parameter, which is shared by a subset of the sensors in the network and needs to be communicated to the rest of the sensor network or to an external controller. This scenario, which we call *medium access with a shared message*, is relevant in several applications pertaining to networked control and coordination, in which the whole network needs to perform an action synergistically. In our case, each of the active sensor nodes has the same piece of information, namely the *shared message*, which they want to deliver to the other nodes in a single shot, i.e., without the possibility of retransmissions or coordination [6]. This is relevant both in ultra-reliable low-latency communication (URLLC), in which the strictest constraint is the latency requirement, and in wide-area scenarios, in which energy consumption is the foremost concern. In this context, traditional throughput maximization-based schemes are extremely inefficient, as they assume each packet carries independent information. We further assume that the active set evolves in a random fashion from one time slot to the next, and each sensor knows only its own membership of the current active set. The challenge here is that a node from the active set cannot coordinate with the other active nodes to send the shared message. The problem looks deceptively simple, but coordinating with limited signaling or shared prior knowledge is extremely difficult. Exploiting correlations in the activity of the sensors to maximize the throughput has been considered in [7], [8]. A related, but different problem has been treated in [9], where the objective is to reliably transmit a shared alarm message by superposing individual signals; this is different from the collision model adopted here, and does not require coordination between the sensors.

We first prove the existence of an optimal deterministic solution to this coordination problem, which is then shown to be NP-hard by modeling the correlations in sensor activity as edge weights in a graph. The coordination problem with deterministic strategies is an instance of a slightly modified weighted graph coloring problem (WGCP) [10], which is a well-known NP-hard problem. Distributed solutions to the WGCP exist [11], and have been used in communications scenarios [12], but they either require signaling between the nodes or more extensive shared feedback. Our scenario allows for extremely limited signaling, as the sensors only receive an acknowledgment (ACK) in case of correct packet reception. We provide a clustering-based solution, which is near-optimal if only two sensors are active at a given time. However, this does not generalize to larger active sets. Instead, we introduce a heuristic to construct the clusters in the general case. We also propose an efficient distributed learning solution using multi-armed bandits (MABs), which can learn correlation patterns and adapt to them without any signaling except for the ACK.

The rest of the letter is organized as follows: Sec. II presents the system model and theoretical analysis. Sec. III presents our two heuristic solutions, which are evaluated through numerical simulations in Sec. IV. Finally, we conclude the paper and present ideas for future work in Sec. V.

Shaan ul Haque (shaanhaque2016@gmail.com) and Siddharth Chandak (chandaks@stanford.edu) are with the Department of Electrical Engineering, Indian Institute of Technology Bombay, India. Siddharth Chandak is also with the Department of Electrical Engineering, Stanford University, USA. Federico Chiariotti (fchi@es.aau.dk) and Petar Popovski (petarp@es.aau.dk) are with the Department of Electronic Systems, Aalborg University, Denmark. Deniz Gündüz (d.gunduz@imperial.ac.uk) is with the Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom. This work was partly supported by the Villum Investigator Grant "WATER" from the Velux Foundation, Denmark.

II. System model

We consider a set $\mathcal{N}$ of N wireless sensors, which share M orthogonal transmission opportunities in time or frequency. In each time slot t, a random subset $\mathcal{A}(t) \subseteq \mathcal{N}$ of sensors, with cardinality $A(t) = |\mathcal{A}(t)|$, become active. The active set is drawn independently at each slot according to probability mass function (PMF) $p_A(\mathcal{A})$. A shared message, e.g., an alarm signal, is to be transmitted by the active sensors to the inactive nodes (which cannot sense the state of the system) or to an external controller. The objective of the active sensors is to deliver the message over the M opportunities, regardless of which sensor it comes from. The challenge lies in the fact that the active sensors have no knowledge of the other sensors in $\mathcal{A}$, and there is no way to explicitly coordinate. The sensors can only agree on a MAC protocol *a priori*.

It is clear that the inactive sensors at each time slot must remain silent. Each active sensor $a \in \mathcal{A}(t)$ must decide on a transmission pattern, called a *move*, expressed as vector $\mathbf{x}_a \in \{0,1\}^M$, where $x_{a,m} = 1$ means that sensor a transmits at the m-th opportunity. We can represent the moves of all the sensors as an $A(t) \times M$ matrix $\mathbf{X}(t)$, with vector $\mathbf{x}_a$ as its a-th row. In the following, we omit the time index t for readability. We consider a simple collision channel, where the condition for transmission success $\xi(\mathbf{X}, \mathcal{A})$ is:

$$\xi(\mathbf{X}, \mathcal{A}) = I\left(\exists m \in \{1, \dots, M\} : \sum_{a \in \mathcal{A}} x_{a,m} = 1\right), \quad (1)$$

where $I(\cdot)$ is the indicator function, equal to 1 if the condition holds, and 0 otherwise. Note that there is a total of 2^M possible moves for each sensor at each time slot. The constrained problem in which each sensor can transmit only once is a special case of this problem, and in general leads to suboptimal solutions, including in some of the cases that we analyze below. The basic rationale for considering multiple transmissions from the same sensor is the same as in irregular repetition slotted ALOHA (IRSA) schemes [13], which exploit the repetitions to improve throughput and reliability. We can define the *strategy* of node a as the PMF $\phi_a(\mathbf{x})$ over the set of possible moves, and represent the strategies of all the sensors in matrix $\Phi \in [0,1]^{N \times 2^M}$, where element $\phi_{n\ell}$ corresponds to the probability of sensor n choosing move ℓ when it is active. We have $\sum_{x=1}^{2^M} \phi_{n,x} = 1$, $\forall n \in \mathcal{N}$. By applying the law of total probability, we get:

$$\mathbb{E}\left[\xi|\Phi\right] = \sum_{\mathcal{A} \in \mathcal{P}(\mathcal{N})} p_A(\mathcal{A}) \sum_{\mathbf{X} \in \{0,1\}^{N \times M}} \xi(\mathbf{X}, \mathcal{A}) \prod_{a \in \mathcal{A}} \phi_a(\mathbf{x}_a), \quad (2)$$

where $\mathcal{P}(\cdot)$ denotes the power set, and $\phi_a(\mathbf{x}_a)$ denotes the probability of sensor a choosing move $\mathbf{x}_a$. We then define our optimization problem, whose solution gives Φ^*, one of the strategies that maximize the expected delivery probability:

$$\Phi^* = \underset{\Phi \in [0,1]^{N \times 2^M}}{\arg\max} \ \mathbb{E}\left[\xi|\Phi\right]. \quad (3)$$

Theorem 1. *At least one of the optimal solutions to the optimization problem is a deterministic strategy; that is,*

$$\exists\ \Phi \in \{0,1\}^{N \times 2^M} : \mathbb{E}\left[\xi|\Phi\right] = \mathbb{E}\left[\xi|\Phi^*\right]. \quad (4)$$

Proof. As the value of ξ is between 0 and 1, its expected value is bounded in the compact interval $[0,1]$, and the $[0,1]^{N \times 2^M}$ region specified by the constraints on Φ is also compact. Accordingly, there exists at least one global maximum.

Assume that in the optimal solution there is at least one sensor with a non-deterministic policy, which we denote as 1. We can then look at all possible moves $\mathbf{x}_1$ and compute the values of the deterministic strategies for sensor 1:

$$\mathbb{E}\left[\xi|(\delta(\mathbf{x}_1); \Phi^*_{-1})\right] = \sum_{\mathcal{A} \in \mathcal{P}(\mathcal{N}) : 1 \notin \mathcal{A}} p_A(\mathcal{A}) \mathbb{E}\left[\xi|\Phi, \mathcal{A}\right] + \sum_{\mathcal{A} \in \mathcal{P}(\mathcal{N}) : 1 \in \mathcal{A}} p_A(\mathcal{A}) \times \sum_{\mathbf{X}_{-1} \in \{0,1\}^{(N-1) \times M}} \xi((\mathbf{x}_1; \mathbf{X}_{-1}), \mathcal{A}) \prod_{a=2}^{A} \phi_a(\mathbf{x}_a), \quad (5)$$

where $\mathbf{X}_{-1}$ is matrix $\mathbf{X}$ without its first row. We can then substitute this into (2):

$$\mathbb{E}\left[\xi|\Phi\right] = \sum_{\mathbf{x}_1 \in \{0,1\}^M} \phi_1(\mathbf{x}_1) \mathbb{E}\left[\xi|(\delta(\mathbf{x}_1); \Phi^*_{-1})\right]. \quad (6)$$

Then, the strategy Φ' that maximizes the expected value:

$$\Phi' = \left(\underset{\mathbf{x}_1 \in \{0,1\}^M}{\arg\max}\ \mathbb{E}\left[\xi|(\delta(\mathbf{x}_1; \Phi^*_{-1})\right]; \Phi_{-1}\right), \quad (7)$$

will be a deterministic one due to the linearity of the objective with respect to $\phi_1(\mathbf{x}_1)$ and the compactness of the simplex. By repeating this operation for all the sensors with non-deterministic strategies, we can find a deterministic solution $\Phi^{(d)}$ that satisfies the condition in (4). □

Hence, we can focus on deterministic strategies over the discrete set of moves $\{0,1\}^{N \times M}$ instead of the continuous probability space $[0,1]^{N \times 2^M}$ without loss of optimality:

$$\mathbf{X}^* = \underset{\mathbf{X} \in \{0,1\}^{N \times M}}{\arg\max}\ \mathbb{E}\left[\xi|\mathbf{X}\right]. \quad (8)$$

The problem is trivial if only one user is active at a time, i.e., if $A(t) = 1, \forall t$: in that case, there is no interference and the trivial solution $\mathbf{x}_{a,m} = 1, \forall a \in \mathcal{A}(t)$ is always successful. The same happens if $M \geq N$, in which case the access to the medium can be entirely orthogonal. However, the problem is extremely complex in the general case.

Theorem 2. *The problem defined in* (8) *is NP-hard if* $A \geq 2,\ \forall\ \mathcal{A} : p_A(\mathcal{A}) > 0$.

Proof. We will prove the NP-hardness of the problem if all the active sets have $A(t) = 2,\ \forall t$, i.e., if exactly two nodes are active at a given time, by showing its equivalence to an instance of the WGCP [14]. WGCP determines the k-coloring of a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$ with minimum weight [10]: it assigns an integer number $c_v \in \{1, \dots, k\}$ to each vertex $v \in \mathcal{V}$. The optimal k-coloring is then the solution to the following weight minimization:

$$\mathbf{c}^* = \underset{\mathbf{c} \in \{1,\dots,k\}^{|\mathcal{V}|}}{\arg\min} \sum_{(u,v) \in \mathcal{E} : c_u = c_v} w_{u,v}. \quad (9)$$

In our case, we can consider a fully connected graph with $\mathcal{V} = \mathcal{N}$. We assign weights equivalent to the probability of two

sensors being active at the same time, i.e., $w_{u,v} = p_A(\{u, v\})$. We can then define the weight minimization as:

$$\mathbf{X}^* = \underset{\mathbf{X}\in\{0,1\}^{N\times M}}{\arg\min} \sum_{u,v\in\mathcal{N},u\neq v} p_A(\{u, v\})\xi(\mathbf{X}, \{u, v\}). \quad (10)$$

As solving the communication problem is equivalent to solving the WGCP, it is NP-hard for $A = 2$. If there are sets with non-zero probability and size $A > 2$, the problem is equivalent to a weighted hypergraph coloring problem (WHCP), which models active sensor sets as weighted edges between two or more nodes and is also NP-hard [15]. Although the problem is equivalent to a WHCP in terms of complexity, the definition is slightly different, as some combination of strategies might result in a successful transmission even if multiple sensors choose the same move, e.g., if two nodes choose to be silent, while the third transmits. The condition is then on ξ, as in (10), and not on the colors (i.e., the strategies) themselves. □

III. Solving the coordination problem

In the following, we present two solutions to this complex problem. The first solution is based on clustering and requires full knowledge of p_A. The second is based on MAB learning, and requires a training period in which the nodes attempt to communicate and learn how to coordinate. Unlike the clustering-based solution, MAB learning can be performed online, as it does not require an oracle view of the network: while knowing exactly which sensors are active at any given time is necessary to estimate p_A, the sensors can independently implement MABs and try all strategies, using acknowledgments for correctly transmitted packets as their only feedback. The clustering-based solution is limited to the case with $A = 2$, and its generalization to larger active sets is non-trivial. Yet, its main advantage is its immediate applicability when the activation distribution p_A is known, with no training period required, as it is derived analytically.

Although finding the optimal strategy is NP-hard, we can always find an optimal solution by brute force, enumerating all N^{2^M} possible strategies. Naturally, the complexity of this solution makes it impractical in most cases, but we can still perform the computation for the small networks that we consider, in order to show the optimality gap.

A. Clustering-based solution

When $A = 2$, we can use the graph representation that we exploited to prove the NP-hardness of the problem to design a clustering-based solution. In this solution, we will group the sensors into 2^M clusters, and assign the same move to all the sensors in the same cluster. We will have a collision if and only if two sensors from the same cluster are active simultaneously. This approach is suboptimal, but the optimality gap is small in the scenarios we considered.

To minimize the collision probability, we employ the probability of two sensors being together in the active set as a cost $d_{u,v} = p_A(\{u, v\})$, where the total cost of a cluster $\mathcal{C}$ is given by the sum of the pairwise costs in the cluster:

$$d(\mathcal{C}) = \sum_{u,v\in\mathcal{C},u<v} p_A(\{u, v\}). \quad (11)$$

Note that the total cost summed over all clusters is equivalent to the failure probability of the scheme if $|\mathcal{A}| = 2$, i.e., only one pair of sensors can be active at any time.

We employ a divisive clustering approach [16], which tries to minimize this total cost at each step in a greedy fashion. This approach significantly outperformed agglomerative and K-means clustering in our experiments. We use a modified version of the DIANA clustering algorithm [17], which starts from a single cluster, then iteratively splits the cluster by dividing the highest-cost cluster, starting from the highest-cost node in it. If $A > 2$, we need to consider nodes one at a time, starting from the highest-probability one and assigning each node to a strategy iteratively. This heuristic is relatively simple and usually has a small optimality gap, but also requires full knowledge of the activation probability matrix.

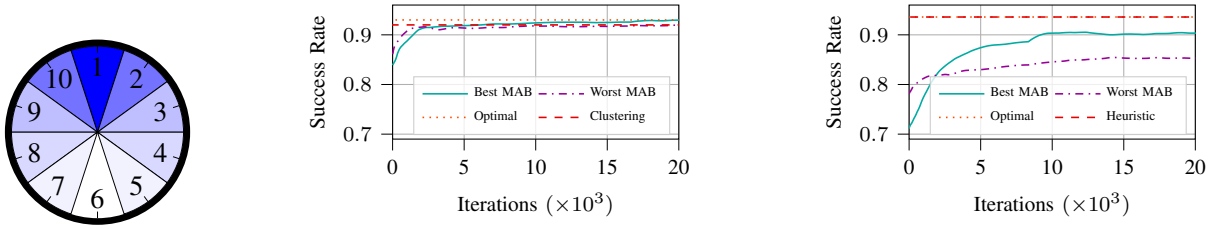
B. Distributed learning approach

It is also possible to learn the optimal policy in a distributed fashion, implementing a MAB for each sensor. MABs are learning agents that have a number of *arms*, which correspond to possible actions, and learn by trying each arm and estimating its expected value over time. There are several sampling strategies to choose which action to use, balancing between exploration (i.e., choosing the action to gain new information) and exploitation (i.e., choosing the action with the highest potential value based on past experience). In our problem, each sensor chooses from 2^M arms, each corresponding to a different move. Each sensor will try different transmission patterns whenever it is active, and get a *reward* $\xi(t) = 1$ if the communication is successful and 0 otherwise, i.e., the shared reward. Since all active sensors are trying to communicate the same piece of information, the communication is considered as successful even if the packet is transmitted by another sensor; that is, the reward depends on the network's success as a whole, and not any individual sensor's action.

Our solution is a distributed version of Thompson sampling for Bernoulli MABs [18]: every time a sensor is active, it chooses a move based on the Thompson sampling algorithm, which is described and explained in [19], and observes the reward. The sampling strategy needs to be entirely distributed, as the only coordination signal available to the agents is the shared binary feedback. The system can also be modeled as a repeated N-player cooperative incomplete information game, in which the unknown information is the membership of the active set. Theorem 1 proves that a pure Nash equilibrium exists, and that it represents the optimal strategy, but there might be other suboptimal equilibria. The Thompson sampling solution converges to an equilibrium with bounded regret [20], in which no agent tries to unilaterally deviate from the joint policy, but the distributed solution might converge to a suboptimal equilibrium, i.e., a local maximum.

The distributed Thompson sampling algorithm is given in Algorithm 1, which is run independently at each sensor: every time a node is active, it uses a Beta distribution to assign probabilities to each action based on the expected reward of the action, then observes the result of the move and updates its values. This requires no coordination, as each sensor acts independently, and leads to quick convergence.

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2022.3181733



(a) Graphical representation. (b) Performance with $A(t) = 2, \forall t$. (c) Performance with $A(t) = 3, \forall t$.

Fig. 1: Results in the regular scenario with $N = 10$, $M = 2$.

Algorithm 1 Distributed Thompson sampling

$\alpha \leftarrow$ ones(2^M)
$k \leftarrow$ones(2^M)
while sampling **do** ▷ Loop over communication rounds
  if active **then** ▷ Only transmit if active
    $X \leftarrow$ChooseAction(M, α, k)
    $\xi \leftarrow$move(X) ▷ Observe result of the action
    $\alpha, k \leftarrow$ThompsonUpdate(X, ξ, α, k) ▷ Update action value
function ThompsonUpdate(X, ξ, α, k)
  $k[X] \leftarrow k[X] + 1$ ▷ Increase sample size for a
  $\alpha[X] \leftarrow \alpha[X] + \xi$ ▷ Update total reward
  return α, k
function ChooseAction(M, α, k)
  $\theta \leftarrow$zeros(2^M)
  for $X \in \{0, \ldots, 2^M - 1\}$ **do**
    $\theta[X] \leftarrow$Beta($\alpha[X], k[X] + 1 - \alpha[X]$) ▷ Sample from Beta distribution
  return $\arg\max(\theta)$

IV. Numerical results

In the following, we show the simulation results for the clustering and learning solutions. We consider a system with N sensor nodes, transmitting over $M = 2$ opportunities. We limit the number of simultaneously active agents to the cases with $A = 2$ and $A = 3$. Although the clustering solution only works in the former, the MAB solution works equally well in both of these cases. Furthermore, we consider two different types of node activation distributions:

1) *Regular activation*: we consider a correlated activation pattern with significant regularity, which might be, e.g., due to the physical location of the sensors. Sensors that are closer together are often simultaneously active, while sensors that are farther apart have a smaller joint activation probability;
2) *General case*: p_A is a general PMF with no apparent regularity. Hand-designing a solution for this case is extremely difficult, as it requires to solve the overall problem and there are no regular features to exploit.

We also tested the algorithm in a case with deterministic activation, i.e., in which each sensor is only in one possible active set. As expected, this case leads to the election of a *leader* in each group, who always transmits, while the other nodes remain silent. The sensors as a whole can transmit 100% of the alarms in this case, and convergence of the distributed MAB to the optimal solution is extremely fast.

A. Regular activation

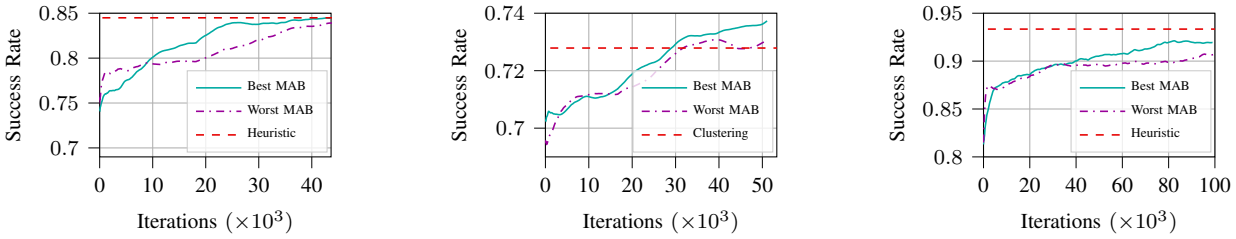
In this model, we assume that $N = 10$ sensors are located around a circle (see Fig. 1a), and that sensors closer to each other are more likely to be active at the same time. At each time t, first sensor in $\mathcal{A}(t)$ is picked at random with probability $\frac{1}{N}$, while the second sensor is picked with probability $p(d)$, which depends on the distance from the first sensor along the circle: if $d = 1$, the probability is 0.275, if $d = 2$, it is 0.125, if $d = 3$, it is 0.075, and if $d = 4$, it is 0.025. This is shown graphically in Fig. 1a: if sensor 1 is picked, the higher probability of picking closer sensors as the second sensor is depicted as a deeper blue. The sensor diametrically opposite to the first one is never picked. When $A = 3$, the third sensor is picked from the same distribution, considering the distance from the second sensor. The strategy is intuitive: nodes that often appear together should have different strategies, so as to avoid collisions and transmit the shared message in at least one of the 2 opportunities. This implies that some sensors are always silent, and others use both opportunities: as this is only a problem in case of two sensors with the same strategy, maximizing the number of different strategies by assigning $(0, 0)$ and $(1, 1)$ to some nodes is the best choice. If $A = 3$, the situation is slightly different, as it is possible to achieve an optimal solution by only using strategies $(1, 0)$ and $(0, 1)$: the objective of the sensors in this case is to avoid a scenario in which all three collide in the same transmission opportunity, as any other combination leads to a successful transmission.

The results are shown in Fig. 1, which includes the best and worst results from 5 independent runs of the distributed MAB approach: in the worst case, this approach might get stuck in a suboptimal equilibrium, corresponding to a local maximum of the reward function. When $A = 2$, the MAB solution always reaches the optimum, while the clustering solution has a small optimality gap. If $A = 3$, the situation is reversed, and the MAB solution converges to a slightly suboptimal solution. In the worst case, the optimality gap of the worst learning curve is close to 0.01. However, the MAB solution does not require any prior knowledge of the activation probabilities, and can be trained online with no knowledge beyond a shared ACK signal in relatively few iterations.

B. General scenario

Finally, we show the results for a general scenario, using a randomly drawn activation probability matrix with $N = 20$. In this scenario, $A = 3$ case is harder, as Fig. 2 shows: the success rate for the best MAB training is slightly below 0.75, while still outperforming the heuristic. If $A = 2$, the MAB and clustering solutions have similar performance, close to 0.85. In this case, we do not know the optimal solution as the brute-force search required to find it is too complex due

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2022.3181733



(a) $M = 2,\ A(t) = 2, \forall t.$ (b) $M = 2,\ A(t) = 3, \forall t.$ (c) $M = 4,\ A(t) = 4, \forall t.$

Fig. 2: Performance in the general scenario with $N = 20$.

to the large number of sensors. However, training times for the MAB solution appear to scale, and the final performance is close to the clustering-based heuristics, which starts with the full knowledge of the environment. The MAB solution with Thompson sampling is general, fast, efficient, and robust, as it can easily deal with errors on the acknowledgments as well as the packets. However, Thompson sampling does not always reach the optimal solution, as it can get stuck in local maxima: using an ε-greedy strategy guarantees better performance after convergence, but requires approximately 1000 times more samples to reach convergence, and as such becomes impracticable in realistic scenarios.

The same happens in a more complex scenario, shown in Fig. 2c, with $M = 4$ and $A = 4$: in this case, the MAB solution performs slightly worse than the heuristic, but it can still reach convergence in relatively few samples, and the optimality gap of the worst learning curve is just 0.03. It is important to note that the constructive heuristic requires full knowledge of the activation probabilities, which are hard to estimate accurately in practice. On the other hand, the MAB solution only requires the ACK signal, and has relatively fast convergence, making it an attractive alternative.

V. Conclusions and Future Work

We have introduced and examined a novel distributed coordination and communication problem, in which multiple wireless sensors need to coordinate and find a common strategy to transmit a shared message. The problem is fundamentally different from general error rate minimization, in which every packet contains different information, and is NP-hard. We proved that a deterministic optimal solution exists, and proposed two heuristics with a small optimality gap in practical conditions. The clustering-based heuristic is close to the optimum, but requires *a priori* knowledge of the system, while the MAB solution can get stuck in a local maximum, but it can be trained online with no additional signaling.

There are several possible avenues of future work on the topic, including the use of more advanced learning mechanisms, such as neural network-based bandits which can generalize experience and converge with fewer training samples. Another interesting research direction is the application of these principles to swarm control, in which the sensors are exchanging information about a shared environment, which they can directly modify by acting in concert. Finally, the scalability of the solutions to scenarios with massive numbers of devices is a significant challenge for future research.

References

[1] J. Cheng, W. Chen, F. Tao, and C.-L. Lin, "Industrial iot in 5g environment towards smart manufacturing," *J. Industrial Inf. Integration*, vol. 10, pp. 10–19, Jun. 2018.

[2] S. Gangakhedkar, H. Cao, A. Ali *et al.*, "Use cases, requirements and challenges of 5G communication for industrial automation," in *Int. Conf. Comm. (ICC) Works.* IEEE, Jun. 2018.

[3] M. Vilgelm, M. Gürsu, and W. Kellerer, "Random access protocols for Industrial Internet of Things: Enablers, challenges, and research directions," in *Wireless Nets. and Industrial IoT*. Springer, Dec. 2020.

[4] X. Jian, Y. Liu, Y. Wei *et al.*, "Random access delay distribution of multichannel slotted ALOHA with its applications for machine type communications," *IEEE Internet of Things J.*, vol. 4, no. 1, pp. 21–28, Sep. 2016.

[5] T. Saadawi and A. Ephremides, "Analysis, stability, and optimization of slotted ALOHA with a finite number of buffered users," *IEEE Trans. Autom. Control*, vol. 26, no. 3, pp. 680–689, Jun. 1981.

[6] Y. Liu, M. Kashef, K. B. Lee *et al.*, "Wireless network design for emerging IIoT applications: Reference framework and use cases," *Proc. IEEE*, vol. 107, no. 6, pp. 1166–1192, Apr. 2019.

[7] A. E. Kalør, O. A. Hanna, and P. Popovski, "Random access schemes in wireless systems with correlated user activity," in *Int. Works. Signal Proc. Adv. in Wireless Comms. (SPAWC)*. IEEE, Jun. 2018.

[8] S. Ali, A. Ferdowsi, W. Saad *et al.*, "Sleeping multi-armed bandit learning for fast uplink grant allocation in machine type communications," *IEEE Trans. Comm.*, vol. 68, no. 8, pp. 5072–5086, Apr. 2020.

[9] K. Stern, A. E. Kalør, B. Soret, and P. Popovski, "Massive random access with common alarm messages," in *Int. Symp. on Info. Theory (ISIT)*. IEEE, 2019.

[10] S.-C. Chang and R. Shrock, "Weighted graph colorings," *J. Stat. Physics*, vol. 138, no. 1, pp. 496–542, Feb. 2010.

[11] L. Barenboim and M. Elkin, *Distributed graph coloring: Fundamentals and recent developments*, ser. Synthesis Lect. Distrib. Comput. Theory. Morgan & Claypool, Jul. 2013, vol. 4.

[12] H. Hernández and C. Blum, "FrogSim: distributed graph coloring in wireless ad hoc networks," *Telecomm. Sys.*, vol. 55, no. 2, pp. 211–223, Feb. 2014.

[13] A. Munari, "Modern random access: An age of information perspective on irregular repetition slotted ALOHA," *IEEE Trans. Comm.*, vol. 69, no. 6, pp. 3572–3585, 2021.

[14] P. Hell and J. Nešetřil, "On the complexity of H-coloring," *J. Combinatorial Theory, Series B*, vol. 48, no. 1, pp. 92–110, Feb. 1990.

[15] I. Dinur, O. Regev, and C. Smyth, "The hardness of 3-uniform hypergraph coloring," *Combinatorica*, vol. 25, no. 5, pp. 519–535, Sep. 2005.

[16] M. Roux, "A comparative study of divisive and agglomerative hierarchical clustering algorithms," *J. Classific.*, vol. 35, no. 2, pp. 345–366, Jul. 2018.

[17] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, Sep. 2009, vol. 344.

[18] G. Ghalme, S. Jain, S. Gujar, and Y. Narahari, "Thompson sampling based mechanisms for stochastic multi-armed bandit problems," in *16th Conf. Auton. Agents and MultiAgent Sys.* ACM, May 2017, pp. 87–95.

[19] D. J. Russo, B. Van Roy, A. Kazerouni *et al.*, "A tutorial on Thompson sampling," *Foundations and Trends in Machine Learning*, vol. 11, no. 1, pp. 1–96, Jul. 2018.

[20] S. M. Asghari, Y. Ouyang, and A. Nayyar, "Regret bounds for decentralized learning in cooperative multi-agent dynamical systems," in *Conf. Uncertainty in Art. Intell. (UAI)*. PMLR, 2020, pp. 121–130.