



HAL
open science

Mitigating User Identification Errors in Resource Optimization for Grant-Free Random Access

Alix Jeannerot, Malcolm Egan, L elio Chetot, Jean-Marie Gorce

► **To cite this version:**

Alix Jeannerot, Malcolm Egan, L elio Chetot, Jean-Marie Gorce. Mitigating User Identification Errors in Resource Optimization for Grant-Free Random Access. IEEE VTC2023-Spring - 97th Vehicular Technology Conference, Jun 2023, Florencia, Italy. hal-04114547

HAL Id: hal-04114547

<https://inria.hal.science/hal-04114547>

Submitted on 2 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche franais ou  trangers, des laboratoires publics ou priv es.



Distributed under a Creative Commons Attribution 4.0 International License

Mitigating User Identification Errors in Resource Optimization for Grant-Free Random Access

Alix Jeannerot
Univ Lyon, INSA Lyon,
Inria, CITI, EA3720
69621 Villeurbanne, France
alix.jeannerot@inria.fr

Malcolm Egan
Univ Lyon, Inria
INSA Lyon, CITI, EA3720
69621 Villeurbanne, France
malcom.egan@inria.fr

Lélio Chetot
Univ Lyon, Inria
INSA Lyon, CITI, EA3720
69621 Villeurbanne, France
lelio.chetot@inria.fr

Jean-Marie Gorce
Univ Lyon, INSA Lyon,
Inria, CITI, EA3720
69621 Villeurbanne, France
jean-marie.gorce@inria.fr

Abstract—In grant-free random access, a key question is how devices should utilize resources without coordination. One standard solution to this problem are strategies where devices randomly select time-slots based on an optimized stochastic allocation rule. However, the optimization of this allocation rule requires accurate knowledge of which devices have been active in previous frames. As user identification algorithms are subject to errors, the expected throughput of the optimized allocation can be highly suboptimal. In this paper, we propose algorithms for optimization of device time-slot allocations that mitigate the impact of user identification errors. We show that when the activity distribution with and without errors is known, then our algorithm converges with probability one to a stationary point. When the activity distributions are not available, we introduce new theoretically-motivated heuristics which significantly improve the expected throughput over existing algorithms and approach the performance when errors are not present.

Index Terms—Grant-free random access, resource allocation, user identification, stochastic gradient descent

I. INTRODUCTION

Resource allocation is a central problem for wireless access networks, particularly for the massive access systems critical for 6G [1], where a large number of devices aim to transmit data to an access point. In 5G-NR and NB-IoT, frames are decomposed into time-frequency resources. A common problem is therefore to determine which devices should be allocated to a given time-slot within a frame.

In order to avoid high levels of control signaling in massive access systems, it is desirable to provide devices with a time-slot selection policy [2], [3], [4], [5]. This means that when a device has data to transmit, limited control signaling is required in order to determine which resources the device should use. The selection of a policy, often a variant of ALOHA (e.g., [2], [6]), is typically based on a criterion such as the expected system throughput, outage probability, or expected sum-rate.

In event-driven IoT systems, devices (e.g., sensors) only require data transmission at irregular frames. The expected throughput therefore depends on the activity distribution, which governs the probability a given device will be active in a frame and can vary device to device. In general, the activity distribution may not be perfectly known, due to imperfect estimation of active devices within each frame. Indeed, the

estimation of active devices, also known as user identification, is currently under active investigation [7], [8].

In this paper, we consider the impact of imperfect estimation of active devices on resource allocation based on the expected throughput. Our focus is on the problem of designing strategies for allocating a single time slot within a frame to an active device [3], [4], [5], relevant for NB-IoT. In this setting, a key challenge is to account for uncertainties in user activity probabilities, which are ignored in most existing work [2], [3], [4], [5] on resource allocation and coding for random access. Our main contributions are as follows:

- We demonstrate that the stochastic resource optimization algorithms recently proposed in [4], [5] can be highly susceptible to user identification errors, leading to suboptimal performance. This is significant as in the absence of user identification errors, these algorithms are optimal in contrast with other heuristic approaches [3] and do not assume perfect knowledge of activity probabilities as in [4], [5].
- To mitigate the impact of the bias arising from user identification errors, we use importance sampling in a manner analogous to sample selection bias correction [9] and bias reduction in private synthetic data [10]. We prove that by weighting the gradient estimates in the algorithms in [4], [5], the resource allocation converges to a stationary point with probability one.
- In practice, the weight for the gradient estimates requires the evaluation of the true activity distribution and the imperfect activity distribution induced by erroneous user identification. As these distributions may be difficult to obtain, we propose two heuristic weights that are readily available in practical systems, leading to new stochastic resource optimization algorithms.
- The proposed algorithms are validated via a numerical study. We show user identification errors can significantly decrease the expected throughput. Realistic scenarios where a low-complexity message passing algorithm [11] is utilized to estimate active devices are also presented. In certain scenarios it is even possible to approach the performance achieved without the presence of user identification errors.

II. SYSTEM MODEL

Consider a time-slotted network consisting of an access point (AP) and N devices equipped with a single antenna and utilizing a common frequency band. Transmissions occur in frames of a fixed duration, where each device n is active with probability p_n in each frame. The activity of each device is denoted by $X_n \sim \text{Ber}(p_n)$ and the joint activity probability mass function by $p_{\mathbf{X}}$. We assume that the activities of the devices are mutually independent. As such,

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{n=1}^N p_n^{x_n} (1 - p_n)^{1-x_n}, \quad \mathbf{x} \in \{0, 1\}^N. \quad (1)$$

In frame t , the activity of all devices is denoted by $\mathbf{X}^t = [X_1^t, \dots, X_N^t]^T$. The activity vector \mathbf{X}^t is also assumed to be mutually independent of $\mathbf{X}^{t'}$, $t' \neq t$.

A. Transmission Protocol

The transmission protocol is detailed in Alg. 1. Each frame consists of a preamble and K data slots. The resource allocation is communicated to devices every f frames to limit downlink usage.

Algorithm 1: Transmission protocol at time t

- 1 (Downlink) Sync signal sent by the AP to indicate beginning of the frame.
 - 2 (Uplink) Active devices send their pilot to the AP during the preamble phase of the frame.
 - 3 (Local at AP) AP detects active devices by estimating \mathbf{X}^t .
 - 4 (Local at device) Each device i randomly selects a slot k_i according to row i of latest allocation matrix \mathbf{A} available at device i .
 - 5 (Uplink) Each active device i sends their data on the selected slot k_i .
 - 6 (Local at AP) AP performs an update of the allocation matrix \mathbf{A}^t based on \mathbf{X}^t .
 - 7 (Downlink) If $t \bmod f = 0$, then the AP sends the new allocation matrix \mathbf{A}^{t+1} to each device.
-

Sync and Preamble: After receiving the sync signal in **Step 1**, each active device in the current frame transmits a pilot sequence of length L during the preamble phase in **Step 2**, denoted by $\mathbf{q}_n \in \mathbb{C}^L$. At the end of the preamble, the AP attempts to identify active devices, as detailed in **Step 3**.

In particular, the received signal $\mathbf{y}^t \in \mathbb{C}^L$ during the preamble phase is given by:

$$\mathbf{y}^t = \sum_{n=1}^N X_n^t h_n^t \mathbf{q}_n + \mathbf{w}_0^t \quad (2)$$

Where $h_n^t \in \mathbb{C}$ is the fading coefficient of the n -th device and $\mathbf{w}_0^t \in \mathbb{C}^L$ is circular complex Gaussian noise with variance τ_w^2 .

An efficient method to estimate the set of active devices in **Step 3** is provided by approximate message passing algorithms (AMP) within the hybrid-gaussian AMP (HGAMP) family [11], [12]. Nevertheless, identification of active devices is not

perfect. Denoting the estimated activity vector at time t by $\hat{\mathbf{X}}^t$, the probability of observing the vector \mathbf{x} at the output of the HGAMP algorithm is denoted by $\Pr(\hat{\mathbf{X}}^t = \mathbf{x})$.

Data Transmission: Each active device transmits in exactly one of the K data slots. In **Step 5**, active devices select a slot randomly, governed by an allocation matrix \mathbf{A}^t with A_{ij}^t , $i \in \{1, \dots, N\}$, $j \in \{1, \dots, K\}$ corresponding to the probability of device i selecting slot j , with $\sum_k A_{nk}^t = 1$, $n = 1, \dots, N$ (as each active user transmits exactly once in a frame).

B. Optimization Objective

The main focus of this paper is the optimization of the slot selection matrix \mathbf{A} . In the following sections, we drop the notation t when the frame index is not relevant. This problem has recently been considered in [3], [4], [5] when the activity vector is perfectly known in each frame. We consider optimization of the expected throughput, defined by [4]:

$$T(\mathbf{A}) = \mathbb{E}_{\mathbf{X}}[T^{\mathbf{X}}(\mathbf{A})], \quad (3)$$

where $T^{\mathbf{X}}(\mathbf{A})$ is defined as

$$T_n(\mathbf{A}; \mathbf{x}) = \sum_{k=1}^K X_n A_{nk} \prod_{\substack{m=1 \\ m \neq n}}^N (1 - X_m A_{mk})$$

$$T^{\mathbf{X}}(\mathbf{A}) = \sum_{n=1}^N T_n(\mathbf{A}; \mathbf{X})$$

The objective in (3) can be interpreted as the expected number of slots in which devices can transmit collision-free.

The resource allocation problem is then to find an allocation matrix \mathbf{A} that maximizes $T(\mathbf{A})$. In particular, we seek a solution \mathbf{A}^* such that

$$\mathbf{A}^* = \arg \max_{\mathbf{A} \in \mathbb{R}_+^{N \times K}; \sum_k A_{n,k} = 1} T(\mathbf{A})$$

Note that the objective $T(\mathbf{A})$ is non-convex in \mathbf{A} . We also highlight that the methods presented in the remaining of this paper are not limited to the maximization of the expected throughput $T(\mathbf{A})$ and are applicable to other objectives; e.g., proportional-fair, expected sum-rate and outage probability [4], [5].

III. THE IMPACT OF IMPERFECT USER IDENTIFICATION

As observed in [4], [5], the problem in (3) can be viewed as a stochastic optimization problem. Under the assumption that the activity vectors $\mathbf{X}^1, \mathbf{X}^2, \dots$ are perfectly known to the AP, a natural solution is stochastic gradient ascent (SGA), which is detailed in Alg. 2. Here,

$$\mathcal{H} = \{\mathbf{A} \in \mathbb{R}_+^{N \times K} : \sum_{k=1}^K A_{n,k} = 1, n = 1, \dots, N\} \quad (4)$$

and $\Pi_{\mathcal{H}}[\cdot]$ denotes the projection to the closest point in \mathcal{H} with respect to the Euclidean norm. The matrix of the

gradient estimate $g(\mathbf{A}; \mathbf{X})$ consists of elements $g(\mathbf{A}; \mathbf{X})_{ql}$, $q \in \{1, \dots, N\}, l \in \{1, \dots, K\}$ defined by [4]:

$$g(\mathbf{A}; \mathbf{X})_{ql} = X_q \prod_{\substack{m=1 \\ m \neq q}}^N (1 - X_m A_{ml}) - \sum_{\substack{n=1 \\ n \neq q}}^N X_q X_n A_{nl} \prod_{\substack{m=1 \\ m \neq n \\ m \neq q}}^N (1 - X_m A_{ml}). \quad (5)$$

Note that in this case, $g(\mathbf{A}; \mathbf{X})$ is an unbiased estimate of $\nabla T(\mathbf{A}) = \mathbb{E}_{\mathbf{X}}[\nabla_{\mathbf{A}} T^{\mathbf{X}}(\mathbf{A})]$ due to the absence of user identification errors.

Algorithm 2: Stochastic optimization algorithm when user identification is error-free.

- 1 Choose initial allocation matrix $\mathbf{A}^1 \in \mathcal{H}$ and step-size sequence $\{\alpha^t\}$ with $\alpha^t > 0$, $t = 1, 2, \dots$
 - 2 $t \leftarrow 1$.
 - 3 **while** *not converged* **do**
 - 4 Based on \mathbf{X}^t , compute an unbiased estimate $g(\mathbf{A}^t; \mathbf{X}^t)$ of $\nabla_{\mathbf{A}^t} T(\mathbf{A}^t)$
 - 5 $\mathbf{A}^{t+1} \leftarrow \Pi_{\mathcal{H}}[\mathbf{A}^t + \alpha^t g(\mathbf{A}^t; \mathbf{X}^t)]$
 - 6 $t \leftarrow t + 1$
 - 7 **end**
-

When the activity vectors are perfectly known, under weak conditions on the step-size sequence (detailed in Theorem 1), the SGA algorithm converges almost surely to a stationary point [4]. In fact, this convergence result holds even when device activity is correlated and the process $\mathbf{X}^1, \mathbf{X}^2, \dots$ is an irreducible Markov process [5, Theorem 1].

However, with realistic user identification algorithms, errors will sometimes be introduced in the estimation of the activity \mathbf{X}^t . As shown in Example III.1, the errors introduced by imperfect user identification can strongly affect the network throughput.

Example III.1. To illustrate the potential impact of user identification errors on resource allocation, consider the simple network consisting of 4 devices sharing two time slots. The true activity probability of the devices is $\mathbf{p} = [0.2 \ 0.3 \ 0.6 \ 0.9]$. Suppose that, due to non-orthogonal pilots or a very low signal-to-noise ratio (SNR), the user identification algorithm might not be able to differentiate between the first and last users. This means that, with probability ϵ , the user activity vector is in fact drawn from $\hat{\mathbf{p}} = [0.9 \ 0.3 \ 0.6 \ 0.2]$. Fig. 1 shows the impact of ϵ on the network throughput $T(\mathbf{A})$ after optimizing \mathbf{A} . The case $\epsilon = 0$ corresponds to the case of perfect user identification and $\epsilon = 1$ to the case where an error event always occurs. Observe that the errors in the user identification can strongly affect the throughput, especially if error events are quite likely to occur (i.e., $\epsilon \approx 1$). We also compare to the throughput of a Frame Slotted Aloha [6] network, i.e., a network that always use a constant matrix of elements $A_{i,j} = \frac{1}{K}$, $\forall i, j$.

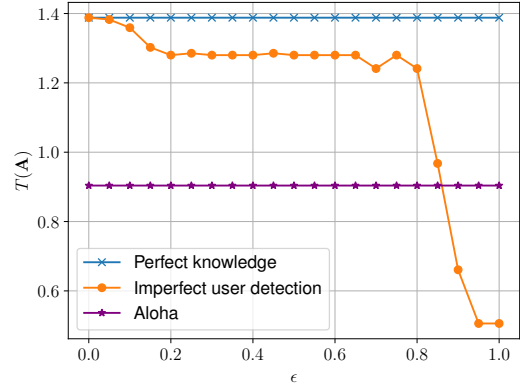


Fig. 1: Expected throughput with user identification errors for a network of $N = 4$ devices sharing $K = 2$ slots. Drawing the sample with probability ϵ from $\hat{\mathbf{p}}$ instead of \mathbf{p} (in orange) might introduce a 60% reduction of the throughput (if $\epsilon = 1$) and lead to a throughput worse than an Aloha network. The blue line corresponds to a perfect identification ($\epsilon = 0$).

IV. A NEW ALGORITHM TO MITIGATE ERRORS

As illustrated in Fig. 1, user identification errors can have a significant impact on resource optimization. In this section, we propose new algorithms in order to mitigate the impact of user identification errors on optimization of the allocation matrix \mathbf{A} .

A. The Effect of Errors

In order to develop a new algorithm to mitigate the impact of user identification errors, we first examine the impact of the errors on the SGA algorithm in Alg. 2. To this end, consider the update rule of Alg. 2:

$$\mathbf{A}^{t+1} = \Pi_{\mathcal{H}}[\mathbf{A}^t + \alpha^t g(\mathbf{A}^t; \mathbf{X}^t)], \quad (6)$$

Observe that the update rule can be rewritten as

$$\mathbf{A}^{t+1} = \mathbf{A}^t + \alpha^t g(\mathbf{A}^t; \mathbf{X}^t) + \alpha^t \mathbf{Z}^t,$$

where $\alpha^t \mathbf{Z}^t$ is the projection term which represent the smallest vector (w.r.t. the Euclidean norm) needed to project $\mathbf{A}^t + \alpha^t g(\mathbf{A}^t; \mathbf{X}^t)$ into the constraint set \mathcal{H} , $g(\mathbf{A}^t; \mathbf{X}^t)$ is the gradient estimate, $\mathbb{E}_{\mathbf{X}}[g(\mathbf{A}^t; \mathbf{X}^t)] = \nabla_{\mathbf{A}^t} T(\mathbf{A}^t) + \beta^t$ with β^t the bias term. We then have

$$\begin{aligned} \mathbf{A}^{t+1} &= \mathbf{A}^t + \alpha^t \left(\mathbb{E}_{\mathbf{X}}[g(\mathbf{A}^t; \mathbf{X}^t)] \right. \\ &\quad \left. + \underbrace{g(\mathbf{A}^t; \mathbf{X}^t) - \mathbb{E}_{\mathbf{X}}[g(\mathbf{A}^t; \mathbf{X}^t)]}_{=\partial \mathbf{M}^t} \right) + \alpha^t \mathbf{Z}^t \quad (7) \\ &= \mathbf{A}^t + \alpha^t (\nabla_{\mathbf{A}^t} T(\mathbf{A}^t) + \beta^t + \partial \mathbf{M}^t) + \alpha^t \mathbf{Z}^t \quad (8) \end{aligned}$$

with $\mathbb{E}[\partial \mathbf{M}^t] = 0$. In the case when there is no error in the user identification, $g(\mathbf{A}^t; \mathbf{X}^t)$ is an unbiased estimate of the gradient $\nabla_{\mathbf{A}} T(\mathbf{A}^t)$, and hence $\beta^t = 0$. This unbiased property is critical in order to guarantee convergence of Alg. 2 [13].

On the other hand, when user identification errors are present, $\beta^t \neq 0$. As a consequence $g(\mathbf{A}^t; \hat{\mathbf{X}}^t)$ is a biased estimate of $\nabla_{\mathbf{A}^t} T(\mathbf{A}^t)$. Hence, there is no guarantee of convergence to a stationary point of (3).

B. A Bias Reduction Method

We have observed that the key cause of the performance reduction in Fig. 1 is the bias in the gradient estimates introduced by imperfect user identification. In order to improve the expected throughput, it is therefore desirable to reduce the bias in $g(\mathbf{A}^t; \mathbf{X}^t)$.

We propose the following solution, inspired by importance sampling [10], [9], to reduce the bias. Consider the weight function $w: \mathbf{x} \mapsto w(\mathbf{x})$ for $\mathbf{x} \in \{0, 1\}^N$ defined by

$$w(\mathbf{x}) = \frac{\Pr(\mathbf{X} = \mathbf{x})}{\Pr(\hat{\mathbf{X}} = \mathbf{x})}, \quad (9)$$

where \mathbf{X} is the true activity vector and $\hat{\mathbf{X}}$ is an estimate; e.g., obtained from HGAMP.

A key property of the importance weight $w(\mathbf{x})$ is that for all $\mathbf{A} \in \mathbb{R}^{N \times M}$ in \mathcal{H} ,

$$\begin{aligned} \mathbb{E}_{\mathbf{X}}[g(\mathbf{A}; \mathbf{X})] &= \sum_{\mathbf{x}} g(\mathbf{A}; \mathbf{x}) \Pr(\mathbf{X} = \mathbf{x}) \\ &= \sum_{\mathbf{x}} g(\mathbf{A}; \mathbf{x}) \frac{\Pr(\mathbf{X} = \mathbf{x})}{\Pr(\hat{\mathbf{X}} = \mathbf{x})} \Pr(\hat{\mathbf{X}} = \mathbf{x}) \\ &= \mathbb{E}_{\hat{\mathbf{X}}}[w(\hat{\mathbf{X}})g(\mathbf{A}; \hat{\mathbf{X}})]. \end{aligned} \quad (10)$$

In other words, $w(\hat{\mathbf{X}})g(\mathbf{A}; \hat{\mathbf{X}})$ is an unbiased estimate of $g(\mathbf{A}; \mathbf{X})$, as long as $w(\mathbf{x}) < \infty$ for all \mathbf{x} such that $\Pr(\mathbf{X} = \mathbf{x}) > 0$. As such, as we will rigorously establish shortly, this choice of weight overcomes the key problem preventing convergence of Alg. 2.

Incorporating the weight defined in (9) into the SGA algorithm yields Alg. 3. This algorithm has the following convergence guarantee.

Algorithm 3: Stochastic optimization algorithm with user identification errors

- 1 Choose initial allocation matrix $\mathbf{A}^1 \in \mathcal{H}$, and step-size sequence $\{\alpha^t\}$ with $\alpha^t > 0$, $t = 1, 2, \dots$
 - 2 $t \leftarrow 1$.
 - 3 **while not converged do**
 - 4 Based on the estimate $\hat{\mathbf{X}}^t$, compute a biased estimate $g(\mathbf{A}^t; \hat{\mathbf{X}}^t)$ of $\nabla_{\mathbf{A}^t} T(\mathbf{A}^t)$.
 - 5 $\mathbf{A}^{t+1} = \Pi_{\mathcal{H}}[\mathbf{A}^t + \alpha^t w(\hat{\mathbf{X}}^t)g(\mathbf{A}^t; \hat{\mathbf{X}}^t)]$
 - 6 $t \leftarrow t + 1$
 - 7 **end**
-

Theorem 1. The iterates \mathbf{A}^t of Algorithm 3, converge almost surely as $t \rightarrow \infty$ to a stationary point provided that the step-size sequence $\{\alpha^t\}$ with $\alpha^t > 0$, $t = 1, 2, \dots$ has the following properties:

$$\sum_{t=1}^{\infty} \alpha^t = \infty, \quad \sum_{t=1}^{\infty} (\alpha^t)^2 < \infty$$

and $w(\mathbf{x}) < \infty$ for all \mathbf{x} such that $\Pr(\mathbf{X} = \mathbf{x}) > 0$.

Proof. Similar to (7), the update rule can be written as

$$\begin{aligned} \mathbf{A}^{t+1} &= \mathbf{A}^t + \alpha^t \left(\mathbb{E}_{\hat{\mathbf{X}}^t}[w(\hat{\mathbf{X}}^t)g(\mathbf{A}^t; \hat{\mathbf{X}}^t)] \right. \\ &\quad \left. + \underbrace{w(\hat{\mathbf{X}}^t)g(\mathbf{A}^t; \hat{\mathbf{X}}^t) - \mathbb{E}_{\hat{\mathbf{X}}^t}[w(\hat{\mathbf{X}}^t)g(\mathbf{A}^t; \hat{\mathbf{X}}^t)]}_{=\partial \mathbf{M}^t} \right) + \alpha^t \mathbf{Z}^t \\ &= \mathbf{A}^t + \alpha^t (\mathbb{E}_{\mathbf{X}^t}[g(\mathbf{A}^t; \mathbf{X}^t)] + \partial \mathbf{M}^t) + \alpha^t \mathbf{Z}^t \end{aligned} \quad (11)$$

$$\begin{aligned} &= \mathbf{A}^t + \alpha^t (\nabla_{\mathbf{A}^t} T(\mathbf{A}^t) + \beta^t + \partial \mathbf{M}^t) + \alpha^t \mathbf{Z}^t \\ &= \mathbf{A}^t + \alpha^t (\nabla_{\mathbf{A}^t} T(\mathbf{A}^t) + \partial \mathbf{M}^t) + \alpha^t \mathbf{Z}^t \end{aligned} \quad (12)$$

where the step (11) is obtained through property:

$$E_{\mathbf{X}}[g(\mathbf{A}; \mathbf{X})] = E_{\hat{\mathbf{X}}}[w(\hat{\mathbf{X}})g(\mathbf{A}; \hat{\mathbf{X}})]. \quad (13)$$

As such, $\beta^t = 0$ as the property in (13) implies that $E_{\mathbf{X}}[g(\mathbf{A}^t; \mathbf{X})]$ is an unbiased estimate of $\nabla_{\mathbf{A}^t} T(\mathbf{A}^t)$.

To establish convergence, we seek to apply Theorem 5.2.1 of [13]. To do so, it is necessary to check the following assumptions:

- A.5.2.1:

$$\sup_t \mathbb{E}_{\hat{\mathbf{X}}^t}[w(\hat{\mathbf{X}}^t)g(\mathbf{A}^t; \hat{\mathbf{X}}^t)] < \infty, \quad (14)$$

which holds provided that $\mathbb{P}(\hat{\mathbf{X}}^t = \mathbf{x}^t) \neq 0$ for all \mathbf{x}^t .

- A.5.2.2: There exists $\bar{g}: \mathbb{R}^{N \times M} \rightarrow \mathbb{R}$ such that

$$\begin{aligned} \mathbb{E}_{\hat{\mathbf{X}}^t}[w(\hat{\mathbf{X}}^t)g(\mathbf{A}^t; \hat{\mathbf{X}}^t)] &= \mathbb{E}_{\hat{\mathbf{X}}^t} \left[w(\hat{\mathbf{X}}^t)g(\mathbf{A}^t; \hat{\mathbf{X}}^t) \right. \\ &\quad \left. | \mathbf{A}^t, w(\hat{\mathbf{X}}^t)g(\mathbf{A}^t; \hat{\mathbf{X}}^t), \forall i < t \right] = \bar{g}(\mathbf{A}^t) + \beta^t \end{aligned}$$

which holds since $\beta^t = 0$ and by setting $\bar{g}(\mathbf{A}^t) = \nabla_{\mathbf{A}^t} T(\mathbf{A}^t)$.

- A.5.2.3: $\nabla_{\mathbf{A}^t} T(\mathbf{A})$ is continuous, which can be immediately verified from the definition of the objective.

- A.5.2.5:

$$\sum_t \alpha^t |\beta^t| < \infty, \quad (15)$$

which holds as $\beta^t = 0 \forall t$.

In addition, a condition on the constraint set \mathcal{H} must be verified. As the same constraint is presented in [4] for the case without errors, we refer the reader to the statement and proof of the constraint condition in [4, Theorem 1]. \square

Theorem 1 shows that even when there are errors, Alg. 3 converges to stationary points of the problems in (3). This is in contrast to the standard SGA algorithm in Alg. 2, for which these guarantees only hold when there are no errors.

C. Proposed Algorithm

While the weight in (13) yields convergence of Alg. 3 to a stationary solution, it must first be computed. In practice this can be challenging for two reasons:

- (i) the true activity distribution may not be perfectly known or is difficult to sample from;
- (ii) the distribution of the estimated activity $\hat{\mathbf{X}}$ must be computed.

We propose the following solution to address (i) and (ii). In order to perform user identification, an estimate of $\Pr(\mathbf{X} = \mathbf{x})$ is required for the prior within the HGAMP algorithm. This prior is typically estimated via the expectation-maximization algorithm coupled with HGAMP [8]. In the computation of the weight in (9), the probability $\Pr(\mathbf{X} = \mathbf{x})$ can therefore be approximated by the prior used within HGAMP.

The HGAMP algorithm also produces estimates of the channels h_n^t . As such, samples of X_n^t and \mathbf{y}^t in (2) can be simulated using the prior estimate of $\Pr(\mathbf{X} = \mathbf{x})$. Let $\epsilon(\mathbf{p})$ be the error probability of HGAMP for a given true activity \mathbf{p} and $e \sim \text{Ber}(\epsilon(\mathbf{p}))$, as a consequence, an estimate $\Pr(\tilde{\mathbf{X}} = \mathbf{x})$ of $\Pr(\hat{\mathbf{X}} = \mathbf{x} | e = 1)$ (the posterior of HGAMP given that an error is made) and $\epsilon(\mathbf{p})$ can be obtained via Monte Carlo estimation. This gives the following approximation for the weighting function:

$$\hat{w}_1(\mathbf{x}) = \frac{\Pr(\mathbf{X} = \mathbf{x})}{(1 - \epsilon(\mathbf{p}))\Pr(\mathbf{X} = \mathbf{x}) + \epsilon(\mathbf{p})\Pr(\tilde{\mathbf{X}} = \mathbf{x})}. \quad (16)$$

Unfortunately, when N is very large, Monte Carlo estimation of $\Pr(\tilde{\mathbf{X}} = \mathbf{x})$ and $\epsilon(\mathbf{p})$ has a high sample complexity. It is therefore desirable to consider further approximations. A convenient choice is

$$\hat{P}(\mathbf{x}) = (1 - \epsilon(\mathbf{p}))\Pr(\mathbf{X} = \mathbf{x}) + \frac{\epsilon(\mathbf{p})}{2^N - 1}. \quad (17)$$

Here, (17) is obtained by considering every possible vector to be equally likely whenever an error is made by HGAMP. An approximation of the weight in (13) is then given by

$$\hat{w}_2(\mathbf{x}) = \frac{\Pr(\mathbf{X} = \mathbf{x})}{\hat{P}(\mathbf{x})}. \quad (18)$$

If the number of users is large, $\epsilon(\mathbf{p})$ can also be approximated to a value close to one (as it will be very likely to miss at least one user).

Accounting for the challenges in computing the weight in (i) and (ii) then leads to Alg. 4 where \hat{w} is either equal to \hat{w}_1 or \hat{w}_2 . Due to the approximations used in the computation of these weights, the bias in the estimated gradients is not zero, as the importance sampling property does not hold anymore, unlike the weight in (5). Nevertheless, as we show in the Sec. V, significant improvements in the performance are obtained over the standard SGA algorithm in Alg. 2, which does not account for the presence of errors in user identification.

Algorithm 4: Stochastic Optimization Algorithm with Approximate Weight Function

- 1 Choose initial allocation matrix $\mathbf{A}^1 \in \mathcal{H}$, and step-size sequence $\{\alpha^t\}$ with $\alpha^t > 0$, $t = 1, 2, \dots$
 - 2 $t \leftarrow 1$.
 - 3 **while** *not converged* **do**
 - 4 Based on the estimate $\hat{\mathbf{X}}^t$, compute a biased estimate $g(\mathbf{A}^t; \hat{\mathbf{X}}^t)$ of $\nabla_{\mathbf{A}^t} T(\mathbf{A}^t)$.
 - 5 $\mathbf{A}^{t+1} = \Pi_{\mathcal{H}}[\mathbf{A}^t + \alpha^t \hat{w}(\hat{\mathbf{X}}^t) g(\mathbf{A}^t; \hat{\mathbf{X}}^t)]$
 - 6 $t \leftarrow t + 1$
 - 7 **end**
-

V. NUMERICAL RESULTS

While we have established that Alg. 3 converges to a stationary point with probability one, this is under the assumption that the weight can be computed with no approximation error. In practice, Alg. 4 must be used with the weight approximated by \hat{w}_1 or \hat{w}_2 . This inevitably leads to bias in gradient estimates and hence a reduction in the final expected throughput. To investigate the reduction compared with the case of perfect knowledge, we perform a numerical study focusing on two scenarios.

In the first scenario, a network consists of $N = 10$ devices and an AP, where each active device transmits on one of $K = 4$ slots. Each device is assumed to transmit independently with probability p_n , with

$$\mathbf{p} = [.19 \ .20 \ .21 \ .23 \ .31 \ .55 \ .71 \ .82 \ .85 \ .94] \quad (19)$$

In the preamble phase, user identification errors are assumed to occur with probability ϵ . In the case an error occurs, the estimated activity vector $\hat{\mathbf{X}}$ is assumed to have independent components with $\Pr(\hat{X}_n = 1) = 1 - p_n$. This error distribution corresponds to a worst case scenario and is therefore of interest to validate that Alg. 4 can effectively mitigate the impact of the errors on slot allocations. The initial matrix \mathbf{A}^1 is randomly selected by setting the elements to $A_{i,j} = \frac{1}{K} + z_{ij}$, with $z_{ij} \sim \mathcal{N}(0, 0.05)$ and then projecting the resulting matrix into \mathcal{H} .

Fig. 2 plots the expected throughput with optimized allocation matrix \mathbf{A} for varying error probability ϵ . Five algorithms are considered: perfect knowledge (Alg. 2) and Aloha (uniform slot allocation for each user, as in III.1) which serve as references; the unweighted case corresponding to Alg. 2 but with a biased estimate of $\nabla_{\mathbf{A}} T(\mathbf{A})$ due to the errors; and two curves corresponding to the proposed Alg. 4 with weights \hat{w}_1 and \hat{w}_2 , respectively. The step-size for each algorithm is set to $\frac{1}{50}$ and the algorithms run for $f = 5000$ iterations. The results are obtained by averaging over 10 runs.

Observe that as expected, the expected throughput with perfect knowledge yields the best performance, while there is significant performance degradation in the unweighted case based on Alg. 2. We highlight that the proposed algorithm in Alg. 4 significantly outperforms the unweighted case for both weights \hat{w}_1 and \hat{w}_2 . For most values of ϵ , Alg 4 is also seen

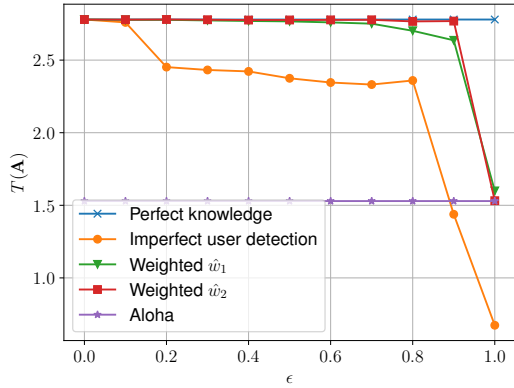


Fig. 2: Comparison of the different algorithms for varying error probability ϵ and activity governed by (19).

to yield performance close to the case of perfect knowledge, which suggests that the impact of user identification errors is largely mitigated. Note also that unless $\epsilon = 1$, all the proposed methods outperform Frame Slotted Aloha.

In the second scenario, Alg. 4 is again compared with the unweighted case, the perfect decoding case and an Aloha system in a network with $N = 10$ devices, $K = 2$ slots and activity vector governed by $\mathbf{p} = [.11 \ .17 \ .18 \ .23 \ .25 \ .34 \ .57 \ .63 \ .70 \ .94]$. In contrast with the first scenario, the distribution of the estimated activity $\Pr(\hat{\mathbf{X}} = \mathbf{x})$ is based on user identification using the HGAMP algorithm which exploits circular complex Gaussian pilot sequences of length $L = 5$. Note that $L < N$ meaning that active users are likely to mixed by the decoder. The weights \hat{w}_1 and \hat{w}_2 are estimated from 3000 prior simulations. The initial matrix is obtained as previously. Each algorithm is run 15 times for $f = 15000$ iterations with a constant step-size of $\alpha^t = \frac{1}{20}$.

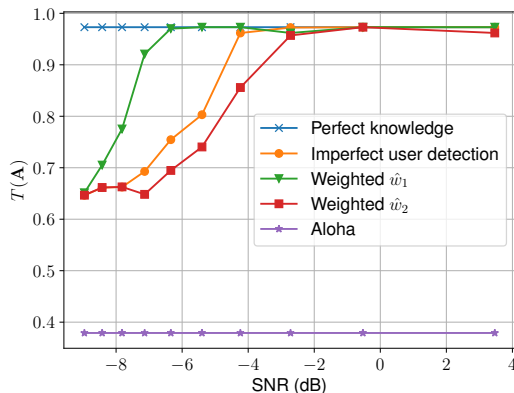


Fig. 3: Expected throughput of the different algorithms as a function of the SNR.

Fig. 3 plots the performance of each algorithm for varying SNR in the preamble phase. For low values of SNR, the

HGAMP algorithm often yields poor user identification and there are limited differences in the performance. On the other hand, as the SNR is increased, while there are still errors it can be observed that the proposed Alg. 4 with weight \hat{w}_1 yields a significant performance improvement over the unweighted case. The reason for the poor performance of weight \hat{w}_2 is that the assumption of uniformly distributed activity in the case of an error is a poor approximation, especially if there are many users. It is interesting to note even when the SNR is small, the algorithms are still outperforming Aloha.

VI. CONCLUSION

Stochastic resource allocation in grant-free random access relies on estimates of device activity via user identification algorithms, such as HGAMP. However, in practice errors are introduced which lead to potentially significant performance reductions. In this paper, we have developed an algorithm that can mitigate the impact of errors via importance sampling. Numerical results show significant performance improvements over existing algorithms, in many cases approaching the performance when user identification errors are not present.

REFERENCES

- [1] Y. L. Lee, D. Qin, L.-C. Wang, and G. Sim, "6G massive radio access networks: Key applications, requirements and challenges," *IEEE Open Journal of Vehicular Technology*, vol. PP, pp. 1–1, 12 2020.
- [2] E. Paolini, G. Liva, and M. Chiani, "Coded slotted ALOHA: a graph-based method for uncoordinated massive access," *IEEE Transactions on Information Theory*, vol. 61, pp. 6815–6832, Dec. 2015.
- [3] A. Kalør, O. Hanna, and P. Popovski, "Random access schemes in wireless systems with correlated user activity," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018.
- [4] C. Zheng, M. Egan, L. Clavier, A. E. Kalør, and P. Popovski, "Stochastic resource optimization of random access for transmitters with correlated activation," *IEEE Communications Letters*, vol. 25, no. 9, pp. 3055–3059, Sep. 2021.
- [5] C. Zheng *et al.*, "Stochastic resource allocation for outage minimization in random access with correlated activation," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2022.
- [6] J. Wieselthier, A. Ephremides, and L. Michaels, "An exact analysis and performance evaluation of framed ALOHA with capture," *IEEE Transactions on Communications*, vol. 37, no. 2, pp. 125–137, 1989.
- [7] J. Ahn, B. Shim, and K. B. Lee, "EP-based joint active user detection and channel estimation for massive machine-type communications," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5178–5189, 2019.
- [8] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Transactions on Signal Processing*, vol. 68, pp. 764–779, 2020.
- [9] B. Zadrozny, "Learning and evaluating classifiers under sample selection bias," *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, vol. 2004, 09 2004.
- [10] S. Ghalebikesabi, H. Wilde, J. Jewson, A. Doucet, S. Vollmer, and C. Holmes, "Bias mitigated learning from differentially private synthetic data: A cautionary tale," *arXiv preprint arXiv:2108.10934*, 2021.
- [11] L. Chetot, M. Egan, and J.-M. Gorce, "Joint identification and channel estimation for fault detection in industrial IoT with correlated sensors," *IEEE Access*, vol. 9, pp. 116692–116701, 2021.
- [12] Q. Zou, H. Zhang, D. Cai, and H. Yang, "Message passing based joint channel and user activity estimation for uplink grant-free massive MIMO systems with low-precision ADCs," *IEEE Signal Processing Letters*, vol. 27, pp. 506–510, 2020.
- [13] H. Kushner and G. G. Yin, *Stochastic approximation and recursive algorithms and applications*. Springer Science & Business Media, 2003, vol. 35.