Abstract—In this paper, we study a multi-agent game between $N$ agents, which solve a consensus problem, and receive state information through a wireless network, that is controlled by a Base station (BS). Due to a hard-bandwidth constraint, the BS can concurrently connect at most $R_d < N$ agents over the network. This causes an intermittency in the agents’ state information, necessitating state estimation based on each agent’s information history. Under standard assumptions on the information structure, we separate each agent’s estimation and control problems. The BS aims to find the optimum scheduling information structure, we separate each agent’s estimation and control problems. The BS aims to find the optimum scheduling information structures that minimizes a weighted age of information based performance metric, subject to the hard-bandwidth constraint. We first relax the hard constraint to a soft update-rate constraint and compute an optimal policy for the relaxed problem by reformulating it into an MDP. This then inspires a sub-optimal policy for the bandwidth constrained problem, which is shown to approach the optimal policy as $N \to \infty$. Next, we solve the consensus problem using the mean-field game framework. By explicitly constructing the mean-field system, we prove the existence of a unique mean-field equilibrium. Consequently, we show that the equilibrium policies obtained constitute an $\epsilon$-Nash equilibrium for the finite-agent system.

Index Terms—Age of information, mean-field games, constrained Markov decision processes, networked control systems.

I. INTRODUCTION

Networked control systems have garnered a lot of research attention over the past decades owing to their appearance in diverse time-critical applications, such as cyber-physical security, real-time monitoring in surveillance systems, autonomous vehicular systems, and Internet-of-Things [2], [3]. to name a few. Often called networked-multi-agent systems, these typically constitute a substantial number of dynamically evolving agents, which are distributed in space and utilize common resources to exchange information remotely [4] between their respective sensors and controllers. This is advantageous in that it allows for decentralized execution of the designated tasks. However, while decentralization reduces the storage complexity at the servers, limited information availability due to bandwidth-constraints can directly affect the control performance of each agent. Thus, appropriate information structures [5] need to be assigned to each system component alongside timely and accurate transmission of timesensitive sensor measurements to the corresponding control units, which is essential in ensuring desired system performance.

In this paper, we consider a discrete-time problem involving $N + 1$ players ($N$ agents and a Base Station) where the $N$ cost-coupled agents are engaged in a consensus building scenario [6], [7]. Each agent constitutes two active decision makers—a controller and an estimator—jointly minimizing a quadratic cost function in a team setting. The agents have access to their state information through a wireless communication network, controlled by a central Base Station (BS). A prototypical diagram depicting the hierarchical structure between the $N + 1$ players is shown in Fig. 1. As in the case of the real-world wireless communication systems, the medium connecting the BS with the agents has a limited bandwidth of $R_d (< N)$ units, which acts as a bottleneck towards uninterrupted transmission of information over the network. Hence, the agents only have partial access to their state information. Thus, the BS must efficiently schedule transmissions so as to minimize the agents’ estimation errors caused by the intermittent communication and also to enable timely communication at the control units. For this purpose, we propose a novel Age of Information (AoI) based metric, called the Weighted-AoI (WAoI), to optimally schedule the wireless communication between the agents. This poses a novel $N + 1$ player game where the $N$ agents are individually trying to solve a consensus problem among themselves while the BS is trying to minimize an aggregate weighted metric (to be precisely defined later) using a scheduling policy. Due to the presence of a large population of agents communicating over the network, we solve the consensus problem within a corresponding mean-field game (MFG) setting. Further, the scheduling problem, where the goal is to minimize the WAoI subject to a bandwidth constraint at the scheduler, is a combinatorics problem, and thus we first solve a relaxed...
problem using a Markov Decision Process (MDP) formulation, and use the solution to construct an asymptotically optimal solution to the original scheduling problem as the agents grow in number.

Related Literature: Ever since the seminal works [8], [9], [10], established no-dual effect and separation properties in systems with imperfect information, a lot of the literature has been concerned with developments on information structures in stochastic decision making problems [5], [12]. Since networked systems involve remote decision making, (un)availability of information at each decision maker can significantly impact the control performance of the system. A number of works have considered networked control problems involving both estimation and control with uninterrupted [13], [14], and intermittent communications [15], [16], [17]. In [15], the authors consider the uplink transmission problem using local schedulers and carry out the co-design of the control and scheduler policies of independent control loops with unconstrained bandwidth. In the current work, however, we consider the scheduling problem over a hard-bandwidth constrained downlink when the agents are cost-coupled, hence, in a game setting. While [16] studies the optimality of an event-based scheduler under constraints on the absolute and average number of resource acquisitions, [17] derives information disclosure policies between an observer and an estimator under a bandwidth constraint over the number of transmissions, in a team setting. Additionally, most earlier works [18, Chs. 6–8] exclude the set of communication instants from the information structure of the decision maker, the inclusion of which while it conveys side information on the current plant state, makes the multi-agent problem, harder to solve. The inclusion of the same is considered in [18, Ch. 2], [19] for a single agent setting, and in [15] for a multi-agent setting, but the overall design considered is only suboptimal. What also differentiates our setting from the ones in [19], [20] is that they consider local schedulers (or event-triggers) within each feedback loop while we consider continuous sensing (at each discrete instant) and a hierarchical structure between agents and the BS as in Fig. 1. Further, the above-mentioned works contain no strategic interaction between the agents as in a game setting, and an additional challenge with the above designs is that of scalability in large population systems such as vehicle-to-everything (V2X) networks, traffic control networks, and industrial Internet-of-Things, to name a few.

To appropriately handle the concerns posed by increasing number of agents in the population, there has been a rapid growth in interest in the MFG setting, since the early works [21], [22], [23], to solve multi-agent problems with strategic interactions. This is due to its efficiency in handling the scalability issue that emerges in finite population games as the number of agents increases. The key idea in MFGs is that, as the number of agents approaches infinity, the effect of individual deviations on equilibrium disappears. This leads to an aggregation effect and the game problem consequently reduces to a stochastic optimal control problem of a generic agent alongside a consistency condition. While the above works consider continuous-time agent dynamics, the discrete-time counterpart is considered in [6], [24] and [25], among others. The linear quadratic mean-field game (LQ-MFG) setting [26] with linear agent dynamics and quadratic cost functions (with consensus terms) serves as a standard, but significant benchmark for general MFGs. Most works in LQ-MFGs consider continuous and reliable communication, with the possible exception of [25] which considers passive (probabilistic) scheduling over unreliable channels. Under suitable assumptions on the probabilities of erasure of actuation and measurement signals, the paper [25] derives approximate-Nash equilibrium strategies for the finite agent game. A recent work [7] deals with a game where the agents are cost-coupled and have access to their state through a noisy intermittent (individual) channel. The problem in this paper, however, differs from the ones in [7], [25] since we actively schedule transmissions over the transmission medium, and extend the unconstrained setting to the case where the agents are connected via a wireless-network structure with a limited bandwidth. We utilize the emerging notion of AoI to devise the optimal scheduling policies.

Age of information was introduced in [27] to measure the timeliness of information in communication networks. After the initial works on AoI which mostly focus on the queuing aspect, scheduling problems in AoI have received significant attention recently. The maximum-age-first policy has been shown to minimize the average AoI for a multi-source system where only one sample can be transmitted at a time with random delay [28] and over a channel with transmission errors [29]. In [30], the authors consider Maximum-age-difference (MAD) policy for preemptive and non-preemptive scheduling policies and numerically observed its performance. In [31], it is shown that the Maximum Weighted Age Reduction (MWAR) policy minimizes a special class of age-penalty functions. The optimality of Whittle’s index policy to minimize AoI has been studied in [32]. Age-optimal scheduling policies are obtained by using MDP formulation in [33], [34], [35], [36], [37].

Recently, AoI has also been studied in the context of networked control problems [38], [39], [40]. In [38], the mean-square estimation (MSE) performance of a system is optimized through HARQ schemes by using MDP formulation. While in [39], the authors compare the performances of AoI based and Value of Information (VoI) based scheduling algorithms, the paper [40] proposes a discounted error scheduler by using MDP approach for a truncated state-space of the AoI. In [41], the authors propose a VoI based event-triggering policy which achieves a lower MSE compared to the AoI-based and periodic updating policies. Surveys [42], [43] provide more detailed literature reviews on AoI. We also note that in contrast to [40], we do not truncate the state space for deriving scheduling policies and the results in our work apply to an unbounded state space, whereby we show by analysis that the optimal policy is characterized as a threshold policy. Finally, we note that a recent paper [2] considers an AoI motivated Internet-of-Things problem within the realm of MFGs and is formulated in the spirit of [44]; however, it

1See [11] for variations of the problem considered in [9].
is fundamentally different from our problem since our focus here is on scheduling policy design from the scheduler to the individual systems and not on the broadcast of information over the uplink.

**Contribution:** The main contributions of this work and its comparison with available literature are as follows.

- We solve a large population game problem involving $N$ linear networked dynamical agents interacting strategically with each other to minimize individual but coupled quadratic cost functions (as in [7], [25]) and a base station (BS) aiming to minimize a performance measure by scheduling communication over a hard-bandwidth constrained wireless medium.

- Inspired by the age of incorrect information (AoII) metric [45], we introduce a weighted AoII metric as the performance measure for the BS, which is a function of the estimation error between the plant state and the controller state (which results due to intermittent information transmission by the BS to the controller) and the AoI at the controller. Such a metric is in contrast to those available in the literature, which either include only the current AoI as the cost [28], [29], [33] to transmit information in a timely fashion or only the Vol to minimize the system uncertainty [39], [41]. Additionally, metrics based on AoII are control unaware and do not involve the presence of a feedback signal. Thus, our metric is novel and appropriately motivated.

- Since the optimal scheduling problem of the BS belongs to the class of multi-armed restless bandits problems [32], finding an optimal policy is difficult. Hence, we propose a suboptimal solution to the scheduling problem. We show that our proposed solution is asymptotically optimal in the limit as $N \to \infty$. This is in contrast to results in the literature which either consider concave or affine running costs for the scheduler [33], [46]. Further, the asymptotic results proposed in [37] are oblivious to the rate of convergence in the stand-alone scheduling problem. On the contrary, in this work, the optimization problem posed at the scheduler is (naturally) non-convex and the proposed solution admits a rate of convergence of the order of $O(1/\sqrt{N})$ for the AoI evolving in an unbounded state space.

- Next, due to the difficulty of dealing with the consensus problem in a large (but not infinite) population setting, we use the mean-field game framework to construct $\epsilon$-Nash strategies for the $N$-agent game problem. Using the scheduling policy constructed above, we prove the existence of a unique mean-field equilibrium (MFE) and also establish its linearity. Furthermore, we prove that $\epsilon = O(1/\sqrt{\min_{\theta \in \Theta} |N_\theta|})$, where $|N_\theta|$ denotes the number of agents of type $\theta \in \Theta$.

- Finally, using numerical simulations, we first verify the asymptotic optimality of the scheduling policy and then show that the controller performance improves when the available bandwidth increases. A comparative study with two existing techniques further demonstrates the effectiveness of the proposed design.

An earlier conference version of this paper, referenced as [1], includes some preliminary results on the special case of the problem considered here, namely, when the scheduling problem of the BS is constrained by an average number of transmissions. We constructed an optimal scheduling policy for the same by minimizing a WAoI metric for the BS. Consequently, by employing this policy, we solved the $N$-agent consensus problem by considering its limiting MFG, computing the MFE, and finally showing its approximate Nash property for the finite-agent game problem. This paper substantially extends these results by considering, among other (as listed above), the problem with hard-bandwidth constraints.

**Organization:** In Section II, we formulate the $N+1$ player game and analyze the centralized scheduling problem of the BS in Section III. The hard bandwidth limit is first relaxed to a soft rate limit, and by using an MDP formulation, we are able to show that a randomized threshold policy is optimal for the relaxed problem. This threshold-based policy then inspires an asymptotically optimal policy for the original bandwidth-limited scheduling problem (Section IV). In Section V, we solve the consensus problem between agents by first considering an MFG (with infinitely many agents) and proving the existence and uniqueness of the mean-field equilibrium. Then, we show that the decentralized control policies obtained by using the MF analysis constitute an $\epsilon$-Nash equilibrium for the finite agent game problem. In Section VI, by using numerical analysis we demonstrate the empirical performance of both the control as well as the scheduling policies along with detailed comparisons. The paper is concluded in Section VII with some major highlights, followed by six appendices.

**Notations:** $\mathbb{Z}^+$ denotes the set of non-negative integers and $\mathbb{N} = \mathbb{Z}^+ \setminus \{0\}$. For a matrix $S$ and vector $x$, $\|x\|_2^2 := x^T S x$. Further, $[N] := \{1, 2, \ldots, N\}$ and $tr(\cdot)$ denotes the trace of its argument. While Euclidean norm for vectors or the induced 2-norm for matrices is denoted by $\| \cdot \|_2$, the notation $\| \cdot \|_F$ denotes the Frobenius norm of its argument. For real functions, $A(\alpha) = O(B(\alpha))$ is equivalent to the statement that there exists $K > 0$ such that $\lim_{\alpha \to \infty} \frac{|A(\alpha)|}{|B(\alpha)|} = K$. The expression $Y_{a,a+k} := \{Y[a], \ldots, Y[a+k]\}$, $k \geq 0$. All empty summations are set to zero, for example, $\sum_{a=1}^{0}() = 0$. For a square symmetric matrix $X$, we use $X \succeq 0$ to indicate that it is positive semi-definite. For any two matrices $X, Y \succeq 0$, $X \succeq Y$ means $X - Y \succeq 0$. Finally, $\mathbb{I}[-\cdot]$ denotes the indicator function of its argument.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We start by formulating the $N+1$ player game, which we express as 1) a consensus problem between $N$ agents, and 2) a centralized scheduling problem of the BS. Then, we will first solve the scheduling problem to calculate a scheduling policy (which couples the estimation processes of the decoders) by solving a WAoI-based optimization problem at the BS. By utilizing this policy we compute the approximate Nash equilibrium policies for the cost-coupled agents.
A. Consensus Problem

We consider a discrete-time $N$-agent game on an infinite horizon, communicating over a wireless network. Each agent’s plant dynamics evolve according to a linear difference equation

$$X^i[k+1] = A(\theta_i)X^i[k] + B(\theta_i)U^i[k] + W^i[k],$$

for timestep $k \in \mathbb{Z}^+$ and agent $i \in [N]$. Here, $X^i[k] \in \mathbb{R}^n$ and $U^i[k] \in \mathbb{R}^m$ are the state and control actions, respectively of agent $i$. The process noise for agent $i$, $W^i[k] \in \mathbb{R}^n$ has zero mean and bounded positive-definite covariance $K_W(\theta_i)$. Further, we assume that $\sup_{1 \leq k \leq M} K_W(\theta_i) \leq K_0$. Agent $i$’s initial state $X^i[0]$ is independent of the process noise and is assumed to have a symmetric density with mean $\nu_{0i}$ and bounded positive-definite covariance matrix $\Sigma_i$. The time-invariant system matrices $A(\theta_i) \in \mathbb{R}^{n \times n}$ and $B(\theta_i) \in \mathbb{R}^{n \times m}$ depend on the agent type.\(^3\) $\theta_i \in \Theta \coloneqq \{\theta_1, \ldots, \theta_L\}$ which is chosen according to the empirical probability mass function $P_N(\theta = \theta_i), \; \theta \in \Theta$. It is further assumed that $|P_N(\theta) - P(\theta)| = O(1/N)$ for all $\theta \in \Theta$, where $P(\theta)$ is the limiting distribution.

We now state the following assumption on information transmission over the network.

**Assumption 1**: We assume that

(i) the wireless links connecting system components are perfect (free of any noise and transmission errors), and

(ii) the BS can send measurements to the corresponding controllers instantaneously.

As a result of Assumption 1, the information can be transmitted from the plant to the controller for its next action without any delay, if the BS decides to send an update for that particular agent. Further, Assumption 1(i) can be relaxed to include, for instance, additive noise channels or erasure channels. This more general model would require additional back-channels from the decoders/estimators to the scheduler to communicate the best estimate of the state in order to entail no-dual effect as in [7]. This no-dual effect property will be instrumental in the design of both scheduling and control policies (to be shown later). Next, we note that in many information updating systems, the required transmission time for the update is usually less than the time unit used to measure the dynamics of a process [34]. Hence, Assumption 1(ii) is justified.\(^4\)

The $N + 1$-player system is shown in Fig. 1. For each agent, its full-state information is relayed to the decoder $D_i$ through a two-hop network (called uplink and downlink) via a centralized BS, which is then communicated to the controller $C_i$ for generating an actuation signal. The downlink capacity is $R_d < N$, which implies that it acts as a bottleneck for transmission of information from the plants to their respective controllers. Then, under Assumption 1, the state of the $i$th plant as observed by the $i$th decoder is

$$\tilde{z}^i[k] = X^i[k] \times [\xi^{d,i}[k] = 1] + \varphi \times [\xi^{d,i}[k] = 0],$$

where $\xi^{d,i}[k] = 1$ denotes that current state information is transmitted to the $i$th decoder (over the downlink) while $\xi^{d,i}[k] = 0$ stands for no transmission (or $\varphi$). Additionally, between transmission times, the decoder calculates the best minimum mean square estimate (MMSE) $Z_i^i[k] = \mathbb{E}[X^i[k] | l^{d,i}[k]],$ based on its information history $l^{d,i}[k] \coloneqq \{\hat{\zeta}_{0:k}, \xi^{d,i}[k], U_{0:k-1}^{i}, Z_{0:k-1}^{i}\}$, which is the conditional expectation of the state given the information structure. The same is then sent to the controller. We adopt the convention that $\tilde{z}[-1], Z[-1], U[-1]$ are all 0 and $W^i[-1] = X^i[0] - Z^i[0]$ for all $i$. Typically, in the game problems as formulated above, the control action of the agent $i$ can depend on other agents’ state and control actions, and hence, the information history of the $i$th controller would be given by $l^{c,i}[k] \coloneqq \{U_{0:k-1}^i, Z_{0:k}^i\}_{i \in [N]}$. Here, $l^{c,i}[s] = \sigma_i(l^{c,i}[s], s = 0, 1, \ldots, k)$, where $\sigma_i$ is the $\sigma$-algebra adapted to its argument and $l^{c,i} \subset l^{c,i}$ is the space of admissible centralized control policies for agent $i$ [7].\(^5\) We also note that no back-channel is required from the controller to the decoder since the control policy (designed later in Section V) can be input a priori into the decoder; and consequently, it can compute the control actions at its own end.

Now, each agent $i$ aims to minimize its infinite-horizon average cost function

\(^3\)While the type of each agent is taken to be fixed in this paper, the analysis can be extended to the case where it is time-varying and evolves according to an underlying finite-state Markov chain, thereby giving rise to a Markov jump-linear system [47].

\(^4\)If we relax this assumption and allow delays in the update transmissions, one needs to keep track of the additional delay in the AoI of each agent, and we leave such a direction as future work. A preliminary result concerning additional delays can be found in [41].

\(^5\)We will later see that the mean-field game setting will naturally lead to the realistic information structure where it will suffice for the control policy of each agent to depend only on the local state of that agent.
\[ J^N_i \left( \pi_i, \pi_{-i} \right) := \limsup_{T \to \infty} \frac{1}{T} E \left[ \sum_{k=0}^{T-1} \| X_i[k] \|_Q + \| U_i[k] \|_R \right], \tag{3} \]

where \( Q(\theta_i) \geq 0 \) and \( R(\theta_i) > 0 \). The coupling between agents enters through the consensus-like term \( \frac{1}{N} \sum_{i=1}^{N} X_i[k] \). The cost function penalizes deviations from the consensus term and large control effort. We define \( \pi_i \) as \((\pi_i^1, \ldots, \pi_i^{N}) \), where \( \pi_i^j \) denotes a control policy for the \( j \)-th agent. Finally, the expectation in (3) is taken with respect to the noise statistics and the initial state distribution. We note that having access to (and keeping track of) the information of other agents in a large population setting is quite difficult, and hence, we will resort to the MFG framework to characterize decentralized control policies whereby decisions are made based on an agent’s local information. This is defined and treated more formally in Section V. We also remark here that the estimator and the controller work together in a team setting to minimize (3), and as we will see later, can be designed independently of each other to lead to a globally optimal design. Finally, the BS acts as a centralized scheduler and schedules transmissions over the downlink using an optimal scheduling policy in order to minimize the WAoI across all agents, as we define below.

B. Centralized Scheduling Problem

Consider the most recently received observations by the controller \( i \) as \( z_i[s_i] \), where \( s_i[k] = \text{sup} \{ s \in \mathbb{Z}^+ : s \leq k, z_i[s] \neq \varphi \} \) denotes the latest transmission time. By definition, the AoI is the time elapsed since the generation time-stamp of the most recent packet at the plant. Thus, the AoI \( \Delta_i[k] \) at the controller \( C_i \) is given as \( \Delta_i[k] = k - s_i[k] \). More precisely, we have that \( \Delta_i[k+1] = (\Delta_i[k] + 1) \times I(\zeta^{d,i}[k] = 0) \). Thus, the scheduling problem with hardbandwidth constraint to be solved by the BS is:

\[ \text{Problem 1 (Control-Aware Constrained AoI Minimization):} \]

\[
\inf_{\gamma \in \Gamma^d} J(\gamma) := \limsup_{T \to \infty} \frac{1}{T} E \left[ \sum_{k=0}^{T-1} \sum_{i=1}^{N} \eta_i[k] \Delta_i[k] \right]
\]

s.t. \( \sum_{i=1}^{N} \zeta^{d,i}[k] \leq R_d, \forall k, \)

where \( \gamma = [\gamma_1, \ldots, \gamma_{d,N}]^T \in \Gamma^d := \{ \gamma^d \} \) is adapted to \( \sigma \{ z_i(s), s = 0, 1, \ldots, k \} \) is an admissible scheduling policy, \( \ell_i[k] \) is the information history of the BS at instant \( k \) and \( \zeta^{d,i}[k] = \gamma^{d,i}(i\ell_i[k]) \) is the scheduling decision. Further, we define \( \eta_i[k] \) as the importance weight associated with agent \( i \) given by \( \eta_i[k] = E [\| e_i[k] \|^2] \), where \( e_i[k] := X_i[k] - Z_i[k] \) is the estimation error between the plant state and the controller state at instant \( k \). Further, \( \eta_i[k] \) can be thought of as the value of information at the agent controller since it represents the loss in the performance of the system when that particular agent is not connected over the downlink. Finally, the expectation is taken over the stochasticity induced by random policies (which, as we will see later provides an optimal solution to the above problem).

III. Centralized Scheduling Problem Analysis

We start this section by computing the best estimate at the decoder. Then, we show that since the system matrices are time-invariant, we can transform the importance weights (which are functions of the errors) into equivalent functions of the AoI for the respective plants. Also, to avoid cluttering notations, we use the shorthand \( A_i := A(\theta_i), B_i := B(\theta_i) \), and \( K_W := K_W(\theta_i) \), unless specified otherwise. We also assume that agent parameters are known to the BS.

Based on the signal to the decoder (2), the decoder computes the best estimate of the state as \( Z_i[k] = X_i[k] + I(\zeta^{d,i}[k] = 1) + E [X_i[k] | I^{d,i}[k], \gamma^{d,i}[k] = 0] \times \| \zeta^{d,i}[k] = 0 \) where \( E [X_i[k] | I^{d,i}[k], \zeta^{d,i}[k] = 0] = A_i E [X_i[k-1] | I^{d,i}[k-1] + B_i U_i[k-1] + \hat{W}_i[k-1], \hat{W}_i[k-1] = E [W_i[k-1] | \zeta^{d,i}[k] = 0] \). We note here that the presence of transmission instants in the conditioning leads to the extra term \( \hat{W}_i[k-1] \). While this conveys additional information on the state of the plant in the absence of any communication between the BS and the decoder, it is typically hard to compute optimally [19]. However, we show that this term equals 0 due to the setup in Fig. 1.

Consider the setup shown in Fig. 2, where we focus on a single agent and additionally add a hypothetical local scheduler (dash-dotted and pink colored box) is inserted between the plant and the BS which connects both if \( \zeta^{d,i}[k] = 1 \).

![Fig. 2. Focused view of a single agent in Fig. 1. Additionally, a hypothetical local scheduler (dash-dotted and pink colored box) is inserted between the plant and the BS which connects both if \( \zeta^{d,i}[k] = 1 \).](image-url)
Lemma 1: The estimation error $e'[k]$ for all agents is independent of the control inputs, and hence there is no dual effect of control [7]. Moreover,

$$e'[k] = \sum_{i=1}^{\Delta[k]} A_i^{[-1]} W'[k - l] \times I[[\zeta^{d, i}[k] = 0]$$

and the covariance of the estimation error can be formulated in terms of the AoI, i.e., $E[\|e'[k]\|^2] = \sum_{i=1}^{\Delta[k]} tr\left(A_i^{[-1]} A_i^{[-1]} I^T W'[k - l]\right) := h(\Delta[k], A_i, K_{Wi})$.

Proof: We note that the decoder output at a non-transmission instant $k$ can be rewritten by using (4) as $Z'[k] = A_i^{[-1]} s'[k] + \sum_{i=1}^{\Delta[k]} A_i^{[-1]} B_i U_i[k - l]$. The result then follows immediately from the definition of $e'[k]$. In addition, using (5), the covariance of the estimation error is obtained as in the statement of the Lemma, which completes the proof.

Thus, Problem 1 can be equivalently written as:

Problem 2 (Control-Aware Constrained AoI Minimization):

$$\inf_{\gamma^d \in \Gamma^d} J^S(\gamma^d) = \limsup_{T \to \infty} \frac{1}{T} E \left[ \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{T-1} h(\Delta'[k], A_i, K_{Wi}) \Delta'[k] \right]$$

s.t. $\sum_{i=1}^{N} \zeta^{d, i}[k] \leq R_d, \forall k$. (6)

We note that Problem 2 does not only depend on the AoI but also on the system parameters of each agent, such as $A_i$ and $K_{Wi}$. This entails that $g(\Delta'[k], A_i, K_{Wi})$, which is nonlinear in the AoI of an agent, appropriately weighs the AoIs of the agents and takes into account the heterogeneity introduced by the agents’ dynamics. Hence, the choice of the running cost $g(\Delta'[k], A_i, K_{Wi})$ is well motivated. We refer to the same as Control-Aware AoI [39].

Remark 1: Note that, at this point, it may be tempting to remove the AoI term, $\Delta'[k]$, from multiplication in the cost function since the error metric $h(\Delta'[k], A_i, K_{Wi})$ is already a function of it. However, we note that if this term is removed, then, in general, the error over an infinite horizon may approach a finite limit (for instance, consider stable agents). This in turn may cause the AoI of those agents to approach infinity since a trigger for information transmission may never be generated for these agents (as will also be clear from Lemma 2 and the proof of Theorem 1 later). This phenomenon is additionally seen from the empirical analysis in [39]. Thus, the use of a weighted metric as in Problem 2 penalizes both the error as well as the AoI and is appropriately justified.

In the sequel, we solve the scheduling problem in Section IV and the consensus problem in Section V. We start by observing that Problem 2 is a combinatorics problem due to the presence of the bandwidth constraint (6), and is difficult to handle. Thus, we first present a relaxed optimization problem involving a time-averaged soft rate constraint on the frequency of transmissions. This entails that more than $R_d$ users are allowed to transmit over the channel as long as the average number of transmissions satisfy the average constraint over the infinite horizon. Consequently, we bring the $N$-agent scheduling problem into a single agent discrete-time MDP to find the optimal scheduling strategies. Next, we define a relaxed version of Problem 2 as follows:

Problem 3 (Relaxed Control-Aware Constrained AoI Minimization):

$$\inf_{\gamma^d \in \Gamma^d} \limsup_{T \to \infty} \frac{1}{T} E \left[ \frac{1}{N} \sum_{k=0}^{T-1} g(\Delta'[k], A_i, K_{Wi}) \right]$$

s.t. $\sum_{k=0}^{T-1} N \sum_{k=0}^{T-1} \zeta^{d, i}[k] \leq R_d$. (7)

Note that the above problem consists of a time-averaged soft rate constraint (7) over the allowed transmissions rather than a bandwidth constraint as in (6). For this problem, we introduce the Lagrangian function as $L(\gamma^d, \lambda) := \limsup_{T \to \infty} \frac{1}{T} E \left[ \frac{1}{N} \sum_{k=0}^{T-1} g(\Delta'[k], A_i, K_{Wi}) + \lambda (\sum_{k=0}^{T-1} \zeta^{d, i}[k] - \frac{N R_d}{T}) \right]$, where $\lambda \geq 0$ is the Lagrange multiplier [48]. Such a multiplier can be thought of as the cost of scheduling for each agent over the channel. Thus, for a fixed $\lambda$, the decoupled single user optimization problem is defined as follows:

Problem 4 (Decoupled Control-Aware Unconstrained AoI minimation):

$$\inf_{\gamma^d \in \Gamma^d} \limsup_{T \to \infty} \frac{1}{T} E \left[ \frac{1}{N} \sum_{k=0}^{T-1} g(\Delta'[k], A_i, K_{Wi}) + \lambda \zeta^{d, i}[k] \right].$$

Since Problem 4 is solved for a single user, we henceforth drop the superscript $i$ until mentioned otherwise.

IV. DECENTRALIZED SCHEDULING PROBLEM

Next, we formulate Problem 4 into a discrete-time controlled MDP M, defined by the quadruplet $M := (S, \mathcal{A}, \mathcal{P}, C)$. The state space $S = \mathbb{Z}^+$ is the set of all possible AoI of the agent and is countably infinite. The action space is $\mathcal{A} = [0, 1]$, where $a = 1$ denotes that the agent is connected over the channel while $a = 0$ stands for no transmission. Note here that $a$ is different from $e'[k]$, which is constrained by the bandwidth limit. The probability transition function $\mathcal{P}$ denotes the evolution of the states of the controlled system. When $a = 0$, we have $\mathcal{P}(\Delta \rightarrow \Delta + 1) = 1$. When $a = 1$, we have $\mathcal{P}(\Delta \rightarrow 0) = 1$. We further note that although the states evolve deterministically, writing them in the form of an MDP will simplify the notation. The one-stage cost $C(\Delta, a) = g(\Delta, A, K_{Wi}) + \lambda a$ denotes the cost incurred when an action $a$ is taken at the state $\Delta$. Next, we formally define the decentralized scheduling problem.

Problem 5 (Decentralized Scheduling Problem):

$$\inf_{\gamma^d \in \Gamma^d} V(\Delta, \gamma^d) := \left\{ \limsup_{T \to \infty} \frac{1}{T} E \left[ \sum_{k=0}^{T-1} C(\Delta[k], a[k]) \right] \right\}.$$ (8)

In the following, we provide the solution to Problem 5.

A. Single-Agent Deterministic Scheduling Policy

We first state the following lemma, which is used in the main theorem characterizing optimality of a threshold policy.
Lemma 2 [49]: The term \( g(Δ[k], A, K_W) \) is an increasing function of \( Δ[k] \) between any two successive transmission instants \( k_1, k_2 \) such that \( k \in [k_1, k_2] \). Further, \( \lim_{Δ[k] \to \infty} g(Δ[k], A, K_W) = \infty. \)

Definition 1: A scheduling policy \( γ^d \) for the MDP M is \( g(Δ) \)-optimal if it minimizes the time-average cost \( V(Δ, γ^d) \).

We first recall that the expectation in (8) is only due to the stochastic nature of the scheduling policy. However, in the next subsection, we characterize a deterministic optimal solution to Problem 5. Thus, we remove the expectation operator in (8) and proceed with further analysis. First, we have the following theorem, a proof sketch for which is provided in Appendix B.

Theorem 1: There exists a \( g(Δ) \)-optimal stationary deterministic policy solving the Decentralized Scheduling Problem 5 for a fixed \( λ \). Further, it has a threshold structure, i.e., there exists a non-negative integer \( τ := (λ, A, K_W) \) such that \( a = 1[Δ ≥ τ] \).

Having established the threshold structure of the \( g(Δ) \)-optimal policy for the single-agent scheduling Problem 5, we can restrict our attention to the finite-state MDP with the state space \( S' = [0, 1, \ldots, τ] \). Then, the one-stage cost for the time-average cost function satisfies the Bellman’s equation given by

\[
V(Δ) + σ^* = \min_{a ∈ \{0, 1\}} \left[ C(Δ, a) + E[V(Δ')|a] \right]
\]

which can be equivalently written as:

\[
V(Δ) + σ^* = \min\{C(Δ, 0) + V(Δ + 1), C(Δ, 1) + V(0)\}, \tag{9}
\]

where \( σ^* \) is the average cost under the \( g(Δ) \)-optimal policy. We also note here that \( σ^* \) is independent of \( Δ \) because the MDP with state space \( S' \) is irreducible.

Next, we calculate an analytical expression for finding \( τ \) as a function of the Lagrange multiplier and the system parameters. We know from Theorem 1 that \( a = 1 \) is optimal at \( Δ = τ \). Then, from (9), we have \( C(τ, 1) + V(0) - σ^* < C(τ, 0) + V(Δ + 1) - σ^* \), which yields \( V(0) < V(τ + 1) \). Further, at \( Δ = τ - 1, a = 0 \) is optimal. Thus, by the same argument, we have that \( V(τ) ≤ λ + V(0) \), which then yields \( V(τ) ≤ λ + V(0) < V(τ + 1) \). Furthermore, by using \( a = 1 \) at \( Δ = τ \), we get

\[
V(Δ) = g(Δ, A, K_W) + λ + V(0) - σ^*. \tag{10}
\]

Since \( V(Δ) \) is monotonically non-decreasing in \( Δ, \exists τ ∈ [0, 1] \) such that \( V(τ + η) = λ + V(0) \), and by using (10), we get that \( σ^* = g(τ + η, A, K_W) \). Further, for \( Δ < τ \), we have \( V(Δ) = σ^* - g(Δ, A, K_W) \), which on summing from \( Δ = 0 \) to \( Δ = τ - 1 \) gives

\[
V(τ) = V(0) + σ^*τ - \sum_{Δ=0}^{τ-1} g(Δ, A, K_W). \tag{11}
\]

Next, we rewrite the expression of \( g(Δ, A, K_W) \) as

\[
g(Δ, A, K_W) = \sum_{l=1}^{Δ} tr\left( (A^{-1})^T A^{l-1} K_W \right) Δ
\]

\[
= \sum_{l=1}^{Δ} tr\left( (A^{-1} K_W)^l \right) Δ
\]

\[
= \sum_{l=1}^{Δ} \left\| A^{-1} K_W^l \right\|_F^2 Δ.
\]

Substituting this in (11), we can calculate the value of \( τ \) by using (10) and (11).

Next, we provide a simplified expression to calculate \( τ \) for scalar systems (\( m = 1 \) in (1)).

1) Scalar Systems: Equating the values of \( V(τ) \) from (10) and (11), we arrive at the equation \( (τ + 1)g(τ + η, A, K_W) - \sum_{l=0}^{τ} g(1, A, K_W) = λ \). By substituting the expression for \( g(\cdot, \cdot, \cdot) \) in this equation yields

\[
\begin{align*}
& f_1(τ, A, K_W, λ) = 0, \quad A \neq 1, \\
& f_2(τ, K_W, λ) = 0, \quad A = 1,
\end{align*}
\]

where \( f_1(\cdot, \cdot, \cdot) = K_W ((τ + 1)(τ + η)A^{2τ+2} - (τ+2)(τ+1)) + \frac{τ^2}{A} - \frac{τ^2}{(A^2-1)} + \frac{τ(τ+1)}{A} - \frac{τ(τ+1)}{(A^2-1)} - λ, \) and \( f_2(\cdot, \cdot, \cdot) = K_W ((τ + 1)(τ + η)^2 - (τ+2)(τ+1)/(τ+1)) - λ. \) We note that (12) is an implicit equation in \( τ \) and \( η \) for given values of \( λ, A \) and \( K_W \). Thus, the value of \( τ \) can be calculated by plotting \( η \) vs \( τ \), and choosing the integer value of \( τ \) for an admissible \( η \).

B. Multi-Agent Randomized Scheduling Policy

In the previous subsection, we showed the existence of a single-agent stationary deterministic policy for a fixed \( λ \). In this subsection, we return to the multi-agent case and obtain the optimal value of \( λ \). Consequently, we use the threshold characterization of the deterministic policy to obtain the optimal solution to Problem 3. The latter policy will be a stationary randomized policy because, in general, a stationary deterministic policy for a constrained optimization problem as in Problem 3 may not exist [50]. Henceforth, we also resume the use of superscript \( i \) to denote the agent index.

Consider the threshold for the \( i^{th} \) agent given by \( τ^i(λ) = \tau^i(λ_i, A_i, K_{Wi}) \). Then the agent is connected to its respective controller at every \( (τ^i(λ) + 1)^{th} \) instant. Or equivalently, its update rate can be given by the quantity \( \frac{1}{τ^i(λ) + 1} \). Thus, under the constraint (7), we have that \( W(λ) := \sum_{i=1}^{N} \frac{1}{τ^i(λ) + 1} \leq R_d \).

In order to find the optimal value of the Lagrange multiplier solving Problem 3, we use the Bisection search procedure [34], which we summarize next. Since \( λ \geq 0 \), we start by initializing two parameters \( λ_l(0) = 0 \) and \( λ_u(0) = 1 \). We then calculate the threshold parameters \( τ^i(λ_l(r)) \) for all \( i \), by using the piece-wise definition in (12). Consequently, we iterate by setting \( λ_l(i+1) = λ_l(i) \) and \( λ_u(i+1) = 2λ_u(i) \) until the constraint on \( W(λ) \) is satisfied for \( λ_l(r) \), for some integer \( r \). Then, we define the interval \( [λ_l(r), λ_u(r)] \). This interval contains the optimal value of the multiplier \( λ^* \), which can be calculated using the Bisection method. The iteration stops when \( |λ_u(r) - λ_l(r)| ≤ \epsilon \), for the iterating index \( m \) and for a suitably chosen \( \epsilon > 0 \).

We next construct the stationary randomized policy solving Problem 3. Define \( λ_l^+(m) = λ_l(0) \) and \( λ_u^+(m) = λ_u(0) \) as obtained above. Further, let the stationary deterministic policies \( τ_d^i(λ_l^+(m)) \) and \( τ_d^i(λ_u^+(m)) \) as obtained from Theorem 1 be those corresponding to \( λ_l^+ \) and \( λ_u^+ \) respectively, where \( τ_d^i(λ_l^+) := \{ τ_d^i(λ_l^+(m)) \}^T \) and \( τ_d^i(λ_u^+) := \{ τ_d^i(λ_u^+(m)) \}^T \). Also, we define \( R_d^i \) and \( R_d^i \) as the total bandwidth used corresponding to...
the multipliers $\lambda^*_i$ and $\lambda^+_i$, respectively. Then, we define the probability $p \coloneqq (R_d - R^0_d)/(R_d - R^\* d)$, and the deterministic policies for all $i$ as:

$$
\gamma_{d1}^{d,i}(\Delta^i) := \begin{cases} 1, \Delta^i \geq t^i_1(\lambda^*_i, A_i, K_{Wi}) \\ 0, \Delta^i < t^i_1(\lambda^*_i, A_i, K_{Wi}) \end{cases}, \quad \text{and}
$$

$$
\gamma_{d2}^{d,i}(\Delta^i) := \begin{cases} 1, \Delta^i \geq t^i_2(\lambda^*_i, A_i, K_{Wi}) \\ 0, \Delta^i < t^i_2(\lambda^*_i, A_i, K_{Wi}) \end{cases}.
$$

Let $\gamma^d_\mathcal{R} := [\gamma^d_1, \ldots, \gamma^d_N]^\top$. The randomized policy $\gamma^d_\mathcal{R}$ for the relaxed Problem 3 can then be obtained as:

$$
\gamma^d_{\mathcal{R}, i} = p \gamma^d_{d1,i} + (1-p) \gamma^d_{d2,i}, \quad \forall i.
$$

Next, we present the following proposition to prove that the randomized policy obtained is indeed optimal for Problem 3.

**Proposition 1 (Optimality of Randomized Policy):** Under Assumption 1, the policy in (13) is optimal for the relaxed minimization Problem 3.

The proof relies on satisfying the assumptions in [50] and using its main results. We omit the details here, which can be found in [49]. Also, as a result of above proof, we will henceforth denote the optimal randomized policy in (13) as $\gamma^d_\mathcal{R} := [\gamma^d_{\mathcal{R}, 1}, \ldots, \gamma^d_{\mathcal{R}, N}]^\top$.

## C. Multi-Agent Hard-Bandwidth Scheduling Policy

In this subsection, we return to our original Problem 2 with the hard bandwidth constraint and construct a new policy satisfying (6). Let $\gamma^d_{i,i}$ be the optimal policy as in (13), and $\alpha^i[k]$ be the scheduling decision for agent $i$ under the relaxed problem. Further, define the set $\Omega[k] := \{i \in [N] \mid \alpha^i[k] = 1\}$, which denotes the agents to be scheduled at time $k$ with the optimal policy, and let $\Omega_k$ be the cardinality of $\omega[k]$. Then, the scheduling decision $\gamma^d_{i,i}[k]$ under the new policy $\gamma^d_{i,i}$, with the hard bandwidth constraint is given as [33]:

- If $\Omega_k \leq R_d$, then $\gamma^d_{i,i}[k] = 1$, $\forall \alpha^i[k] = 1$.
- If $\Omega_k > R_d$, then $\gamma^d_{i,i}[k] = 1$ for a subset $\Omega_k[k] \subset \omega[k]$ of the agents, which are selected at random (with uniform probability) by the BS such that $\Omega_k[k] = R_d$, where $\Omega_k$ denotes the cardinality of $\Omega_k$. The agents in the set $\omega[k] \setminus \omega_k[k]$ remain unselected.

In the following, we show that the costs under $\gamma_{i,i}$ and $\gamma^d$ approach each other asymptotically (as $N \to \infty$), where $\gamma^d := [\gamma^d_{i,i}, \ldots, \gamma^d_{i,i}]^\top$. To this end, we start by defining a new policy $\hat{\gamma}^d$, which transmits exactly the same agents as the relaxed policy, except that, for each additional agent that was not supposed to be transmitted by the hard-bandwidth policy $\gamma^d$, it adds an additional penalty to the cost defined as $\omega(\Delta, A, K_{Wi}) = \sum_{i=1}^{\infty}(1 - R^0_d)g(\tau + i, A, K_{Wi})/(1 - R^0_d) > 0$ if $\Delta \geq \tau$. Further, we let $\{\hat{\Lambda}[1], \hat{\Lambda}[2], \ldots, \hat{\Lambda}[k], \ldots\}$ and $\{\Lambda[1], \Lambda[2], \ldots, \Lambda[k], \ldots\}$ be the sequences of AoIs of the $p$th agent under $\gamma^d_{i,i}$ and $\gamma^d_{i,i}$ (or equivalently $\hat{\gamma}^d_{i,i}$), respectively. Then, it is easy to see that $\omega(\hat{\Lambda}[k], A_i, K_{Wi}) \geq \mathbb{E}[\hat{\Lambda}[k], A_i, K_{Wi}]]$, $\forall k, i$. Let $\gamma^d_{i,i}$ be an optimal policy that solves Problem 2. Then, it immediately follows that

$$
\frac{J^S_{\gamma^d}}{\gamma^d_{\mathcal{R}} \leq J^S_{\gamma^d_{i,i}} \leq J^S_{\gamma^d} \leq J^S_{\hat{\gamma}^d}.
$$

Consider next the Markov chain induced by the relaxed policy $\gamma^d_{\mathcal{R}}$ for the $i$th agent as

$$
\Delta^i[k+1] = \begin{cases} \Delta^i[k] + 1, \text{ w.p. } 1, \quad \Delta^i[k] < \lambda_{ij}^i \\ \Delta^i[k] + 1, \text{ w.p. } 1 - p, \quad \Delta^i[k] = \lambda_{ij}^i \\ 0, \quad \text{ w.p. } p, \quad \Delta^i[k] = \lambda_{ij}^i \\ 0, \quad \text{ w.p. } 1, \quad \Delta^i[k] = \lambda_{ij}^i. \end{cases}
$$

Let $\pi^*[k] := [\pi^*[0[k], \ldots, \pi^*[1(\lambda_1^i)]^\top]$ denote the probability distribution of the states in the above mentioned Markov chain such that $\pi^*[k] \approx \mathbb{P}(\Delta^i[k] = \lambda_{ij}^i), \forall \mathcal{R}$. Define $\pi^* := [\pi^*[0], \ldots, \pi^*[1(\lambda_1^i)]^\top$. Then, since each state in the set $\Omega^d_{i,i} := \{0, 1, 2, \ldots, \tau^d(\lambda_1^i)\}$ is reachable from every other state, the Markov chain in (15) is irreducible. This implies that it admits a unique stationary distribution $\pi^*$ [51]. We now state the following assumption and then present the main result of this section.

**Assumption 2:** The inequality $0 < \|A(\theta)\|_F < \sqrt{1/(1 - \alpha)}$ holds $\forall \theta \in \Theta$, where $\alpha = R_d^\mathcal{R}$.

We remark that Assumption 2 entails finite scheduling cost under the hard-bandwidth policy, and further insights on it are presented later (in Remark 2).

**Theorem 2:** Suppose Assumption 2 holds. Then, given a fixed value of $\alpha$, the deviation of the relaxed scheduling policy $\gamma^d_{\mathcal{R}}$ from the policy $\gamma^d$ is of the order of $O\left(\frac{1}{\sqrt{N}}\right)$. Consequently, as $N \to \infty$, $\gamma^d$ is asymptotically optimal for Problem 2.

**Proof:** The proof is provided in Appendix C. Since the solution to the scheduling problem is complete, in the next section, we proceed to establish the $\epsilon$–Nash property of the mean-field solution.

## V. N–AGENT CONSENSUS PROBLEM

In this section, we solve the second part of the $N + 1$–player game problem, namely, the consensus problem using the central scheduler’s policy. To this end, we first consider the limiting game called the mean-field game (MFG) (as $N \to \infty$). Under this setting, the empirical coupling term in (3) is approximated by a known deterministic sequence (or the MF trajectory) and the closeness of the approximation is justified later in the analysis. This principle is well known in the literature as the Nash certainty equivalence principle [21] and reduces the game problem to a stochastic optimal control problem of a generic agent with a coupled consistency condition. The equilibrium solution obtained (called the mean-field equilibrium (MFE)) will be shown to constitute an $\epsilon$–Nash approximation to the finite agent game problem.

### A. Stochastic Optimal Tracking Control

Consider a generic agent of type $\theta$ from an infinite population with dynamics

$$
X[k+1] = A(\theta)X[k] + B(\theta)U[k] + W[k],
$$

where timestep $k \in \mathbb{Z}^+$, $X[k] \in \mathbb{R}^d$ and $U[k] \in \mathbb{R}^m$ are the state vector and the control input, respectively. Further, the initial state $X[0]$ is assumed to have a symmetric density
function with $E[X[0]] = v_{θ,0}$ and $cov(X[0]) = Σ$, is bounded. Next, $W_k ∈ ℝ^a$ is a zero-mean i.i.d. Gaussian noise with finite covariance $K_W(θ)$. All covariance matrices are assumed to be positive-definite. The objective function of the generic agent is given by

$$J(μ, \bar{X}) := \limsup_{T→∞} \frac{1}{T} \mathbb{E} \left[ \sum_{k=0}^{T-1} \|X[k] - \bar{X}[k]\|^2_{Q(θ)} + \|U[k]\|^2_{R(θ)} \right], \quad (17)$$

where $μ := (μ[1], μ[2], \ldots) ∈ M_{d,con}$ and is an admissible control policy of the generic agent. Further, the admissible set $M_{d,con} := \{μ | μ$ is adapted to $α(I_{d,con}[s], s = 0, 1, \ldots)\}$ is the space of decentralized control policies for the generic agent and $I_{d,con}[0] := [z[0]], I_{d,con}[k] := \{U_{0,k-1}, Z_{0,k}\}, k ≥ 1$, is the local information history of the generic agent. Recall that this is in contrast to $I_{con}[k]$ (the centralized history at the controller), which includes information of other agents as well. This implies that $M_{d,con} ⊆ M_{r,con}$. The information structure for the generic agent’s decoder is defined similar to that in Section II-A, except that the superscript $i$ is removed. Further, $\bar{X} = (X[1], X[2], \ldots)$ is the MF trajectory and denotes the infinite player approximation to the consensus term in (3) and serves to decouple the otherwise coupled game problem into a linear-quadratic tracking (LQT) problem via introducing indistinguishability between agents. Finally, the expectation above is taken with respect to the noise statistics and the initial state distribution.

To solve the LQT problem with dynamics in (16) and the cost in (17), we first state the following assumptions.

**Assumption 3:** (i) The pair $A(θ), B(θ)$ is controllable and $(A(θ), √Q(θ))$ is observable.

(ii) The MF trajectory belongs to the space of bounded functions, i.e., $\bar{X} ∈ \mathcal{X} := \{\bar{X}[k] ∈ ℝ^n | \|\bar{X}\|_\infty := \sup_{k≥0} \|\bar{X}[k]\| < \infty\}$.

We remark here that Assumption 3 is quite standard in the MF-LQG literature [24], [25]. Next, we define a MFE by introducing the following operators: 1) $Ψ : \mathcal{X} → M_{d,con}$, defined as $Ψ(\bar{X}) = \text{argmin}_{μ ∈ M_{d,con}} J(μ, \bar{X})$, which outputs the optimal policy for a given MF trajectory, and 2) $Λ : M_{d,con} → \mathcal{X}$, also called the consistency operator, that regenerates a MF trajectory for a control policy as obtained in 1) above.

**Definition 2 (Mean-Field Equilibrium (MFE) [24]):** The pair $(μ^*, \bar{X}^*) \in M_{d,con} × \mathcal{X}$ is a MFE if, $μ^* = Ψ(\bar{X}^*)$ and $\bar{X}^* = Λ(μ^*)$. In other words, $\bar{X}^* = Λ(Ψ(\bar{X}^*))$.

Now, the central scheduling policy is fixed from the previous section, and similar to [7] we have the following result for the optimal tracking control of a generic agent.

**Proposition 2:** Consider the generic agent (16) with the controller state as in (4) and the cost function in (17). Let $\tilde{α} := 1 - α$. Then, under Assumptions 1-3, the following hold true:

1) The optimal decentralized control action of the generic agent of type $θ$ is given as

$$U^*[k] = -\Pi(θ)Z[k] - L(θ)r[k+1], \quad (18)$$

where $L(θ) = \left(R(θ) + B(θ)^\top K(θ)B(θ)^{-1}B(θ)^\top\right)\Pi = L(θ)K(θ)Λ(θ)$, and $K(θ)$ is the unique positive-definite solution to the algebraic Riccati equation $K(θ) = A(θ)\left[K(θ)A(θ) - K(θ)B(θ)\Pi(θ) + Q(θ)\right]$.

Further, the trajectory $r[k]$ satisfies the backward dynamics $r[k] = H(θ)^\top r[k+1] - Q(θ)\bar{X}[k]$, with the initial condition $r[0] = -\sum_{j=0}^{∞} (H(θ)^\top)^j Q(θ)\bar{X}[j]$ and $H(θ) = A(θ) - B(θ)\Pi(θ)$ is Hurwitz.

2) The difference equation for $r[k]$ above has a unique solution in $X$, given as

$$r[k] = -\sum_{j=k}^{∞} \left(H(θ)^j\right)^\top Q(θ)\bar{X}[j]. \quad (19)$$

3) The optimal cost is bounded above as:

$$J(μ^*, \bar{X}) ≤ \limsup_{T→∞} \frac{1}{T} \mathbb{E} \left[ \sum_{k=0}^{T-1} \bar{X}[k]^\top Q(θ)\bar{X}[k] \right.$$

$$\left. - r[k+1]B(θ)L(θ)r[k+1] + \left|A(θ)^\top K(θ)B(θ)\Pi(θ)\right| \times \sum_{n=1}^{T/2} \sum_{\tau=1}^{n} \left[A(θ)^{n-\tau} - A(θ)^{n-\tau}K(θ)W(θ)\right] \right.$$

$$\left. + \frac{\left|K(θ)W(θ)\right|_{F}}{\left|A(θ)\right|_F} \times \frac{\left|\left[A(θ)^{n-\tau}\right]_{F} - 1 - \left|A(θ)\right|_F^2\tilde{α}\right|}{\tilde{α}} \right]. \quad (20)$$

**Proof:** The proof is provided in Appendix D.

**Remark 2:** We remark here that as seen from the proof of the above proposition, while Assumption 2 entails a finite optimal cost as in (20), it warrants that one requires a higher downlink bandwidth ($R_d$) to accommodate a higher degree of instability among the agents. This can be seen from the fact that $ρ(Λ(θ)) ≤ \|Λ(θ)\|_F$, where $ρ(·)$ denotes the spectral radius of the argument matrix. Thus, the bound of the Frobenius norm as in Assumption 2 in turn bounds the maximum eigenvalue that can be stabilized using the optimal control in (18). This can be further observed from the special case of a scalar system, where the Frobenius norm equals the magnitude of the only eigenvalue, and is bounded as a function of the available bandwidth.

### B. Mean-Field Analysis

In this subsection, we prove the existence and uniqueness of the MFE by explicitly constructing the MF operator as follows.

Consider the closed-loop system in (4) with the control policy in (18) as

$$Z[k+1] = \begin{cases} H(θ)Z[k] - B(θ)L(θ)r[k+1] + W[k], \quad \xi^d[k+1] = 1, \\ H(θ)Z[k] - B(θ)L(θ)r[k+1], \quad \xi^d[k+1] = 0, \end{cases}$$

where $ξ^d[k]$ is chosen from the hard-bandwidth policy $γ^d$ of Section IV-C. The above can be rewritten as

$$X[k+1] = H(θ)X[k] - B(θ)L(θ)r[k+1] + B(θ)\Pi(θ)e[k] + W[k],$$
where $e[k]$ is defined in (5). By taking expectation and substituting $r[k]$ from (19), we get
\[
\hat{X}_0[k] = H(\theta)^k v_{0,0} + \sum_{j=0}^{k-1} H(\theta)^{k-j-1} B(\theta) L(\theta) \times \sum_{s=j+1}^{\infty} (H(\theta)^{s-j-1})^T Q(\theta) \hat{X}[s]
\] (21)
where $\hat{X}_0[k] = E[X[k]]$ is the aggregate dynamics across agents of type $\theta$ and we use the fact that $E[e[k]] = 0$. As a result, the MFE (if it exists), is invariant under the choice of the scheduling policy as long as we can compute an optimal control of the generic agent with bounded control cost, as shown in Proposition 2.

Next, using the empirical distribution from Section II, we define the MF operator as
\[
\mathcal{T}(\hat{X})[k] := \sum_{\theta \in \Theta} \hat{X}_0[k] \mathcal{P}(\theta).
\] (22)
Then, we state the following assumption before we prove the main result.

**Assumption 4:** It holds that $\hat{H}(\theta) := \|H(\theta)\| + \nu < 1, \forall \theta \in \Theta$, where $\nu = \sum_{\theta \in \Theta} \|Q(\theta)\| \frac{\|B(\theta)L(\theta)\|}{(1-\|H(\theta)\|)^2} \mathcal{P}(\theta)$.

It is quite common in the literature [23], [24] to invoke the above assumption; although it is stronger than the corresponding assumption in [6], [25], it leads to the linearity property of the MF trajectory, which can then be computed offline.

**Theorem 3:** Under Assumptions 2-4, the operator $\mathcal{T}(\hat{X}) \in \mathcal{X}$, $\forall \hat{X} \in \mathcal{X}$, and there exists a unique $\hat{X}^* \in \mathcal{X}$ such that $\mathcal{T}(\hat{X}^*) = \hat{X}^*$. Furthermore, $\hat{X}^*[k]$ follows linear dynamics, i.e., $E[\hat{E}^*] \in \mathcal{E} = \{\hat{E} \in \mathbb{R}^{1 \times n} : \|\hat{E}\| < 1\}$, $\hat{X}^*[k+1] = \hat{E}\hat{X}^*[k]$, where $\hat{X}^*[k]$ is the equilibrium MF trajectory of the agents, and $\hat{X}^*[0] = \sum_{\theta \in \Theta} v_{0,0} \mathcal{P}(\theta)$.

The proof relies on proving contraction properties of operators in a suitable norm and then using the Banach’s fixed point theorem. A more detailed proof can be found in [49].

**Remark 3:** We note that Theorem 3 gives us a unique MFE and a feedback control law which is linear in the agent’s state and the equilibrium trajectory. This makes the computation of this trajectory tractable, which would otherwise have involved a non-causal infinite sum. Further, as a result of the linear MF trajectory, we will (later) be able to compute an explicit convergence rate between the equilibrium trajectory and the coupling term in (3) in Proposition 3.

Next, we prove that the equilibrium policy obtained above constitutes an $\epsilon$–Nash equilibrium for the $N$–agent system (1) with the cost (3).

**C. $\epsilon$–Nash Analysis**

In this subsection, we show that the decentralized equilibrium control policy obtained from the MF analysis is approximately Nash for the finite-agent system.

We start by proving the following Lemma, which shows that the closed-loop system is stable under the equilibrium MF control policy (18), i.e.,
\[
\sup \max_{N \geq 1} \limsup_{T \to \infty} \frac{1}{T} E \left[ \sum_{k=0}^{T-1} \|X^s,i[k]\|^2 \right] < \infty.
\]

**Proof:** The proof is provided in Appendix E.

Next, we state the following proposition, which establishes that the equilibrium MF trajectory approximates the finite-agent state average in the mean-square sense.

**Proposition 3:** Suppose that Assumptions 1-4 hold. Then, the equilibrium MF trajectory converges in the mean-square sense to the coupling term in (3) with a rate of $O\left(\frac{1}{\min_{\epsilon>0} |N_\epsilon|}\right)$, where $N_\epsilon \subset |N|$ is the subset of agents of type $\theta$, given its cardinality $|N_\epsilon| > 0, \forall \epsilon \in \Theta$. More precisely, we have that $\epsilon_T(N) = O\left(\frac{1}{\min_{\epsilon>0} |N_\epsilon|}\right)$, where $\epsilon_T(N) = \frac{1}{T} \sum_{k=0}^{T-1} E\left[ \frac{1}{N} \sum_{j=1}^{N} X^s,j[k] - \hat{X}^*[k] \right]^2$.

**Proof:** The proof is provided in Appendix F.

Next, we prove the $\epsilon$–Nash property of the equilibrium control policy. To this end, we introduce the following metrics. Also, we suppress the arguments of $Q$ and $R$ for brevity. Let
\[
J_i^N(\mu^{\epsilon,i}, \mu^{\epsilon,-i}) = \limsup_{T \to \infty} \frac{1}{T} E \left[ \sum_{k=0}^{T-1} \|X^s,i[k]\|^2 - 1 \frac{\sum_{j=1}^{N} X^s,j[k]\|^2}{N} + \|U^s,i[k]\|^2 \right]
\] (23)
be the cost of agent $i$ when all agents follow the policy under the MFE;
\[
J_i^N(\mu^{\epsilon,i}, \hat{X}^*) = \limsup_{T \to \infty} \frac{1}{T} E \left[ \sum_{k=0}^{T-1} \|X^s,i[k] - \hat{X}^*[k]\|^2 + \|U^s,i[k]\|^2 \right]
\] (24)
be the cost of agent $i$ with the consensus term replaced by the equilibrium MF trajectory; and
\[
J_i^N(\pi^{\epsilon,i}_i, \mu^{\epsilon,-i}) = \limsup_{T \to \infty} \frac{1}{T} E \left[ \sum_{k=0}^{T-1} \|X^i,\pi^{\epsilon}_i[k]\|^2 - 1 \frac{\sum_{j=1}^{N} X^j[k]\|^2}{N} + \|V^i[k]\|^2 \right]
\]
be the cost of agent $i$ when it deviates from the MF policy while all other agents follow the MF policy. Here, $X^i,\pi^{\epsilon}_i[k]$ is the state of agent $i$ at time $k$ when it chooses a control policy $\pi^{\epsilon}_i \in \mathcal{M}^i_{\epsilon,con}$. Furthermore, the control action $V^i[k]$ is derived from $\pi^{\epsilon}_i$. To make the argument and the result more concrete, we now have the following definition:

**Definition 3** [7]: The set of control policies $\{\mu^{\epsilon,i}, 1 \leq i \leq N\}$ constitute an $\epsilon$–Nash equilibrium with respect to the cost functions $J_i^N, 1 \leq i \leq N$, if there exists $\epsilon > 0$, such that $J_i^N(\mu^{\epsilon,i}, \mu^{\epsilon,-i}) \leq \inf_{\mu^{\epsilon,-i} \in \mathcal{M}^i_{\epsilon,con}} J_i^N(\pi^{\epsilon,i}_i, \mu^{\epsilon,-i}) + \epsilon$, $\forall i \in [N]$.
an $\epsilon$-Nash equilibrium for the $N$-agent LQ-mean field game with bandwidth limits. In particular, we have that

$$J_N^2(\pi^*, \mu^*\big|\cdot) = \inf_{\pi \in \mathcal{M}_2} J_N^2(\pi, \mu^*\big|\cdot)$$

and

$$= O\left(\frac{1}{\min_{\theta < \Theta} |N\theta|}\right).$$

Next, we present some simulations to empirically validate the theoretical results.

**VI. SIMULATIONS**

In this section, we provide an empirical analysis of the theoretical results. We simulate an $N$-agent system with scalar dynamics. We first point out a key observation which may seem not obvious. Consider the set of $A$ and $K_W$ values for 7 agents as $A = \{0.1, 0.3, 0.5, 0.7, 1.0, 1.3, 1.4, 1.5\}$ and $K_W = \{3.0, 5.0, 1.0, 2.0, 4.0, 0.1, 2.0\}$. It may be tempting to think that the values for the $\tau_i$’s would be lower for the unstable agents than for the stable ones. However, the values that are obtained are $\tau = \{1, 0, 1, 0, 2, 1\}$ and $\tau_u = \{1, 0, 1, 1, 1, 2, 1\}$ with $p = 0.33$ for $R_d = 4$. This shows, for instance, that agent 6 is not connected over the network for a longer period than a more stable agent 5. This should be expected because $E[\|e'[k]\|^2]$ for all $i$ depends not only on $A_i$ but also on $K_W$. Thus, a higher $\tau_u^0$ (or $\tau_u^0$) is explained by a low value of $K_W$.

Next, Fig. 3(a) shows a variation of $\tau_u^0$ and $R_d$ for $N = 6$. For convenience, we plot only $\tau_u^0$ and $\tau_u^0$ can be plotted similarly. The agent parameters used are $A = \{0.1, 0.299, 0.498, 0.697, 0.896, 1.095\}$ and $K_W = \{0.3, 0.9, 1.5, 2.5, 4, 4.5\}$. When $\alpha = \frac{R_d}{N}$ is closer to 1, the values of the threshold parameters $\tau_u^0$ are lower which means that fresh information is transmitted to the agents more frequently, thereby leading to performance improvement. It can be seen that with more restricted bandwidth constraints, the agents are not connected to their respective controllers for longer intervals. Further, it can also be seen from the figure that for small values of $R_d$, stability of a system becomes more important, while for higher $R_d$ values, the noise covariance becomes a dominating factor. This observation is aligned with intuition and can also be inferred from the equations for the error covariance and the cost function.

Next, we empirically demonstrate the asymptotic optimality of the hard-bandwidth policy (Fig. 3(b)) as proved in Theorem 2. The purpose of Fig. 3(b) is two-fold. First, we show the behavior of WAoI under: 1) the relaxed policy (of Section IV-B), and 2) the hard-bandwidth policy (of Section IV-C) with a bandwidth of $R_d = 0.6N$. We plot the average WAoI of both these policies from $N = 5$ up to $N = 1500$, and a simulation time of $T = 100000$ seconds, and we observe that the WAoI for both asymptotically approach each other as $N$ increases. This means that as the number of agents increases, the approximation error between the bandwidth-constrained and the soft-constrained problems decreases to zero. Second, in the same figure, we also compare our proposed approach with two existing policies, namely, the maximum-age-first (MAF) [29], [39] and the estimation error-based scheduling policy (EES) [39]. Clearly, the proposed policy minimizing the WAoI performs better, which shows that neither MAF nor EES is optimal for the considered case. Further, both latter policies attain a constant suboptimality margin for any possible value of $N$. Finally, we empirically evaluate the behavior of $N = 900$ agents under the MFE policy (18) in Figures 4 and 5, and further compare it with the MAF and EES policies. We consider three types of agents with $A = 0.5, 1.0, 1.15$, $K_W = 5$ and a horizon $T = 500$ seconds. In Fig. 4, we plot the aggregate control cost (23) per agent as a function of the proportion of available downlink bandwidth. The figure shows a box plot depicting the median (red line) and spread (box) of the average cost per agent over 100 runs for each value of $\alpha$. First, we observe that EES and the proposed policy perform significantly better in the low bandwidth region ($\alpha = 0.25$). This is because the MAF policy transmits information of agents in a round-robin manner without caring about the degree of instability in the end-users’ dynamics while both other policies take those into account. Further, the proposed policy performs better than the EES policy for $\alpha = 0.45$ while for other bandwidth ratios the performances are similar. A similar trend can be also be seen

![Figure 3](image-url)
from Fig. 5, where we plot the average cost per agent of only the most unstable agents with $A = 1.15$, in which case the MAF policy performs even worse. Additionally, another point to note from both figures is that the average cost decreases as the available bandwidth increases as aligned with intuition. The above observations reinstate the theme of this paper that when the end-users have dynamically evolving states (and may typically be heterogeneous), we need to appropriately take into account the agent parameters, which is precisely the motivation behind proposing a control-aware WAOI metric.

**VII. Conclusion & Discussions**

In this paper, we have studied an $N$ agent game problem, where each agent receives state information over a bandwidth-constrained network, which is regulated by a BS. To efficiently schedule agents over the network, we have proposed a WAOI-based scheduling problem for the BS. To solve the former, we have first solved a relaxed version of the problem by proposing a stationary randomized optimal scheduling policy. Then, we have shown the $\epsilon$–Nash property of the equilibrium solution for the finite-agent system. Finally, using numerical simulations, we have first verified the asymptotic optimality of the scheduling policy and then showed that the controller performance improves when the available bandwidth increases. A comparative study with two existing techniques further demonstrated the effectiveness of the proposed design.

Our results can be extended in different directions. The first one is to incorporate imperfections in the channel models. In the current work, we considered deterministic downlink and ideal uplink, which could be extended to include erasure, noise, and/or transmission delays. In such cases, we may still be able to analytically determine the optimal policy at the BS by utilizing the theory of partially observable MDPs and track the AoIs of the agents based on whether the information was received by an agent or not. However, an additional complication may arise because of the fact that the transmitted state of the agents now may not be the actual state but only an estimate inferred at the BS based on the information history of the BS and the signal received over the uplink at the current instant, due to partial observability introduced by the uplink non-idealities.

Another direction in which our results could be extended would be to relax the assumption that agent parameters
are known to the BS, which would lead to the setup of reinforcement learning. In this case, one could either infer the parameters using system identification procedures and work with the identified parameters, or use model free techniques to learn an optimal scheduling policy at the BS without having to know these parameters. Finally, in this work, we have considered the linear system dynamics as in (1). Extension of our results to nonlinear system dynamics would also constitute promising research direction. Here, we also note that in the consensus problem, the coupling among agents can be extended to include more complex interactions through both the dynamics and the costs, the agent dynamics can be extended to include measurement dynamics.

**APPENDIX A**

In this section we provide the proofs of some of the theorems and propositions. For full versions of the proofs, we refer the reader to supplementary materials and also in [49].

**APPENDIX B**

**PROOF OF THEOREM 1**

We provide a proof sketch here. First, we can introduce an $\alpha$-discounted cost function $V_\alpha(\Delta, \gamma^d) := \sum_{k=0}^{\infty} \alpha^k C(\Delta[k], a[k])$, on an infinite horizon, with $0 < \alpha < 1$ and $\Delta[0] = \Delta$. In addition, we can define a scheduling policy $\gamma^d$ as $\alpha$-optimal if it minimizes the total $\alpha$-discounted cost $V_{\alpha}(\Delta, \gamma^d)$ and its associated cost as $V_{\alpha}(\Delta) = \inf_{\gamma^d} V_\alpha(\Delta, \gamma^d)$. Then, we can write down its discounted-Bellman equation $V_\alpha(\Delta) = \min_{a \in \{0, 1\}} \{C(\Delta, a) + \alpha E[V_\alpha(\Delta')]\}$ and can prove the following proposition using the monotonicity property of $V_\alpha(\Delta)$.

**Proposition 4:** There exists an $\alpha$-optimal stationary deterministic policy for a fixed $\lambda$. Further, it has a threshold structure, i.e., there exists a non-negative integer $\tau := \tau(\lambda, A, K_W)$ such that $a_0(\Delta) = 1 - \tau$, where $a_0$ denotes the action under an $\alpha$-optimal policy.

Finally, to complete the proof of Theorem 1, we can make use of the main result of [52], by satisfying Assumptions 1, 2 and 3* in [52]. Then, by considering a sequence of $\alpha_l$ such that $\lim_{l \to \infty} \alpha_l = 1$, the optimal policy $a_{\alpha_l}$, which is $\alpha_l$-optimal from Proposition 4, converges to the $g(\Delta)$-optimal policy [52], which verifies the threshold structure of the latter. Further, the time-average cost $\sigma^* = \lim_{T \to \infty} (1 - a) V_\alpha(\Delta)$ is finite and independent of the initial state [52]. Then, this completes the proof of the theorem.

**APPENDIX C**

**PROOF OF THEOREM 2**

Consider the following:

$$J^S_{\gamma^d} - J^S_{\gamma^d,*} \leq \frac{M}{\sqrt{N\alpha}} \lim_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \sum_{j=1}^{N} \left[ \frac{1}{\sqrt{N}} - \frac{1}{\sqrt{N}} \right] \left[ \frac{1}{\sqrt{N}} - \frac{1}{\sqrt{N}} \right]$$

where $J^S_{\gamma^d}$ and $J^S_{\gamma^d,*}$ are the costs under policies $\gamma^d$ and $\gamma^d,*$, respectively. The expression $g(\cdot, \cdot, \cdot)_s$ denotes the WAoI under the policy $s$. The second equality follows since sample paths of the AoI under $\gamma^d$ coincide with those under the policy $\gamma^d,*$ by definition. The first inequality follows since $\{\cdot\}^+ \leq 1$, and the second inequality follows as a result of Assumption 2 and $M$ being a constant, dependent on system parameters, $\alpha$, and threshold values. The third inequality follows since $\{\cdot\}^+ \leq 1$. Further, in the third inequality, we have that $E[\Omega] = \lim_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \sum_{j=1}^{N} \gamma^d_{R,j}(\Delta[k], a[k])$, which follows as a result of the Ergodic theorem [51] due to irreducibility of the Markov chain (15) and $\gamma^d_{R,j}(\Delta[k], a[k])$ being the indicator of the transmission of agent $j$ in state $i$ under the relaxed policy. The notation MAD($\alpha_l$) stands for the mean absolute deviation of its argument. Note also that $\Omega_k = \sum_{j=1}^{N} \gamma^d_{R,j}(\Delta[k], a[k])$, where $\gamma^d_{R,j}(\Delta[k], a[k])$ are independent binary random variables due to the fact that the relaxed policy was obtained by decoupling the problem into single agent problems. Let $\sigma^l := \left\{ \sum_{j \in S_0} \gamma^d_{R,j}(\Delta[k], a[k]) \right\}$ denote the probability of transmission of agent $i$ under $\gamma^d,*$ in state $j$ under the stationary distribution $\pi^t$. Then, the random variable $\gamma^d_{R,j}(\Delta[k], a[k])$ has zero mean and unit variance. Continuing from (26), we get

$$J^S_{\gamma^d} - J^S_{\gamma^d,*} \leq \frac{M}{\sqrt{N\alpha}} \lim_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \sum_{j=1}^{N} \left[ \frac{1}{\sqrt{N}} - \frac{1}{\sqrt{N}} \right] \left[ \frac{1}{\sqrt{N}} - \frac{1}{\sqrt{N}} \right]$$

where the second inequality is true by Jensen’s inequality and we use the fact that $a(1 - a) \leq 1/4$ for $0 \leq a \leq 1$. Finally, using (14), we have that $J^S_{\gamma^d} - J^S_{\gamma^d,*} \leq J^S_{\gamma^d} - J^S_{\gamma^d,*}$. Thus, as $N \to \infty$, from (14), we have that the hard-bandwidth policy $\gamma^d$ is asymptotically optimal for Problem 2 with order $O(1/\sqrt{N})$. The proof is thus complete.
\( J(\mu^*, \tilde{X}) \leq \text{tr}(K(\theta)K_W(\theta)) + \limsup_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \tilde{X}[k]^{\top}Q(\theta)\tilde{X}[k] \)

\[
= -r[k + 1]^{\top}B(\theta)L(\theta)r[k + 1] + \limsup_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \left\| A(\theta)^{\top}K(\theta)^{\top}B(\theta)\Pi(\theta)\right\|_F \mathbb{E}\left[\|e[k]\|^2\right].
\]

\[ (27) \]

Let \( \tilde{\alpha} := 1 - \alpha \). Then, consider the following:

\[
\mathbb{E}\left[\|e[k]\|^2\right] \leq \sum_{n=1}^{n_{\epsilon}} \sum_{r=1}^{m} tr(A(\theta)^{\top}\tilde{\alpha}^{-1}A(\theta)^{\top}K_{\theta}(\theta))
\]

\[
+ \sum_{n=r_{\epsilon}+1}^{n_{\epsilon}} \sum_{r=1}^{m} tr(A(\theta)^{\top}\tilde{\alpha}^{-1}A(\theta)^{\top}K_{\theta}(\theta)) \tilde{\alpha}^{n-r_{\epsilon}}
\]

\[
\leq \sum_{n=1}^{n_{\epsilon}} \sum_{r=1}^{m} tr(A(\theta)^{\top}\tilde{\alpha}^{-1}A(\theta)^{\top}K_{\theta}(\theta))
\]

\[
+ \left\| K_{\theta}(\theta) \right\|_F \frac{\tilde{\alpha}^{n_{\epsilon}}}{\tilde{\alpha} - 1} \left( \left\| A(\theta)^{\top}\tilde{\alpha}^{-1}A(\theta)^{\top}K_{\theta}(\theta) \right\|_F^2 \tilde{\alpha}^{n_{\epsilon}} \right) - \frac{\tilde{\alpha}^{n_{\epsilon}}}{\tilde{\alpha} - 1}
\]

\[ (28) \]

where \( n_{\epsilon} := n_{\epsilon}(\lambda^*) \), the first inequality follows using the hardbandwidth policy of Section IV-C, and the second inequality follows using Assumption 2, and the fact that \( \|AB\| \leq \|A\|_F\|B\|_F \). Then, combining (27) and (28), we arrive at (20). This completes the proof.

### Appendix E

**Proof of Lemma 3**

We follow the proof technique of [25]. Before proceeding, we drop the * from the superscripts for ease of notation. Then, by using (16) and (18), we can write the closed-loop system as

\[
X[k + 1] = H(\theta)X[k] + B(\theta)\Pi(\theta)e[k] - B(\theta)L(\theta)r[k + 1] + W[k].
\]

Then, using (29), we arrive at

\[
\mathbb{E}\left[\|X[k]\|^2\right] \leq 4\mathbb{E}\left[\|H(\theta)X[0]\|^2\right]
\]

\[
+ 4\mathbb{E}\left[\left\| \sum_{p=0}^{k-1} H(\theta)k^{p-1}B(\theta)\Pi(\theta)e[p]\right\|^2\right]
\]

\[
+ 4\mathbb{E}\left[\left\| \sum_{p=0}^{k-1} H(\theta)k^{p-1}B(\theta)L(\theta)r[p + 1]\right\|^2\right]
\]

\[
+ 4\mathbb{E}\left[\left\| \sum_{p=0}^{k-1} H(\theta)k^{p-1}W[p]\right\|^2\right],
\]

\[ (30) \]

where we used the fact that \( \| \sum_{n=1}^{n_{\epsilon}} x_i \|_2 \leq n \| x_i \|_2 \). Next, the terms on the RHS can be bounded using Proposition 2 and Assumptions 2-3, to arrive at

\[
\limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\sum_{k=0}^{T-1} \|X[k]\|^2\right] \leq 4\theta(\theta)tr(\Sigma_e + v_{\theta_i, 0}v_{\theta_i, 0}^{\top}/(1 - \varepsilon(\theta))) + 4\beta(\theta)Q(\theta)\|B(\theta)\Pi(\theta)\|^2 \times
\]

\[
\left[ \frac{1}{1 - \varepsilon(\theta)} + \frac{1}{(1 - \sqrt{\varepsilon(\theta)})^2} \right] + \frac{\varepsilon(\theta)tr(\Sigma_e)}{1 - \varepsilon(\theta)} + \frac{4\theta(\theta)tr(K_{\theta}(\theta))}{1 - \varepsilon(\theta)},
\]

where \( \beta(\theta) = \sum_{n=1}^{n_{\epsilon}} \sum_{r=1}^{m} tr(A(\theta)^{\top}\tilde{\alpha}^{-1}A(\theta)^{\top}K_{\theta}(\theta)) + \|K_{\theta}(\theta)\|_F \left( \left\| A(\theta)^{\top}\tilde{\alpha}^{-1}A(\theta)^{\top}K_{\theta}(\theta) \right\|_F \right)^2 \left( 1 - \frac{1}{\|\tilde{\alpha}^{-1}\|_2} \right) - \frac{1}{\|\tilde{\alpha}^{-1}\|_2} \right), M_{r_i} := \|r_i\|_\infty, \) and \( \varepsilon(\theta) \geq 1 \) and \( \theta(\theta) \in (0, 1) \) are constants.

Finally, since \( \theta \) is a finite set, we have the desired result.
Lemma 3, we have that there exist $O(\epsilon^2)$ obtained using the definition of $\epsilon^2(N)$ and the fact that 
$$\frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E} \left[ \left\| Y^i_k[k] - \hat{X}^i_k[k] \right\|^2 \right]$$
and $N = \epsilon \theta$, we can deduce $\epsilon(N)$ and $O(1/N)$, we can deduce $\epsilon(N) = O(1/\min_{\theta \in \Theta} |N|)$. The proof is thus complete.

**APPENDIX G**

**PROOF SKETCH OF THEOREM 4**

We can show using Proposition 2 and Cauchy-Schwarz inequality that 
$$\limsup_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E} \left[ \left\| Y^i_k[k] - \hat{X}^i_k[k] \right\|^2 \right] \leq \limsup_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E} \left[ \left\| Y^i_k[k] - \hat{X}^i_k[k] \right\|^2 \right]$$
and $J^i_N(\mu^i, \mu^{-i}) - J(\mu^i, \hat{X}^i)$.

**REFERENCES**

[1] S. Aggarwal, M. A. U. Zaman, M. Bostanci, and T. Başar, "Large population games with timely scheduling over constrained networks," in *Proc. IEEE ACC*, 2023, pp. 4772–4778.

[2] H. Zhang, Y. Kang, S. Han, and H. V. Poor, "Age of information minimization for grant-free non-orthogonal massive access using mean-field games," *IEEE Trans. Autom. Control*, vol. 69, no. 11, pp. 7806–7820, Nov. 2021.

[3] H. Sandberg, S. Amin, and K. H. Johansson, "Cyberphysical security in networked control systems: An introduction to the issue," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 20–23, Feb. 2015.

[4] X.-M. Zhang, Q.-L. Han, and X. Yu, "Survey on recent advances in networked control systems," *IEEE Trans. Ind. Informat.*, vol. 12, no. 5, pp. 1740–1752, Oct. 2016.

[5] A. Malipolopoulos, "On team decision problems with nonclassical information structures," *IEEE Trans. Autom. Control*, vol. 68, no. 7, pp. 3915–3930, Jul. 2023.

[6] M. A. U. Zaman, K. Zhang, E. Miehling, and T. Başar, "Approximate equilibrium computation for discrete-time linear-quadratic mean-field games," in *Proc. IEEE ACC*, 2020, pp. 333–339.

[7] S. Aggarwal, M. A. U. Zaman, and T. Başar, "Linear quadratic mean-field games with communication constraints," in *Proc. IEEE ACC*, 2022, pp. 1323–1329.
Muhammad Aneeq uz Zaman received the B.S. degree in mechatronics engineering from the National University of Science and Technology, Pakistan, and the M.Sc. degree in mechanical engineering from the University of Illinois Urbana–Champaign, IL, USA, where he is currently pursuing the Ph.D. degree with the Coordinated Science Laboratory. His research interests include multiagent reinforcement learning using the mean-field game paradigm among others.

Melih Bastopcu (Member, IEEE) received the B.S. degree in electrical and electronics engineering from Bilkent University, Turkey, in 2016, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland at College Park, MD, USA, in 2020 and 2021, respectively. He is currently a Postdoctoral Research Associate with the Coordinated Science Laboratory, University of Illinois Urbana–Champaign. His research focus is on age of information, resource allocation, low latency systems, communication network design, dissemination of misinformation, and social networks.

Tamer Başar (Life Fellow, IEEE) received the B.S.E.E. from Robert College, Istanbul, and the M.S., M.Phil., and Ph.D. degrees from Yale University. Since 1981, he has been with the University of Illinois Urbana–Champaign (UIUC), where he is currently the Swanlund Endowed Chair Emeritus and the Center for Advanced Study Professor Emeritus of Electrical and Computer Engineering. He is also with the Coordinated Science Laboratory, Information Trust Institute, and Mechanical Science and Engineering, UIUC. At UIUC, he has also served as the Director of CAS from 2014 to 2020, the Interim Dean of Engineering in 2018, and the Interim Director of the Beckman Institute from 2008 to 2010. He has over 1000 publications in systems, control, communications, optimization, networks, and dynamic games, including books on non-cooperative dynamic game theory, robust control, network security, wireless and communication networks, and stochastic network control. His current research interests include stochastic teams, games, and networks; risk-sensitive estimation and control; mean-field game theory; multiagent systems and learning; data-driven distributed optimization; epidemics modeling and control over networks; strategic information transmission, spread of disinformation, and deception; security and trust; energy systems; and cyber–physical systems. He received the Wilbur Cross Medal from Yale University in 2021. In 2021, he has served as the President of IEEE Control Systems Society (CSS), the International Society of Dynamic Games (ISDG), and the American Automatic Control Council (AACC). He has received several awards and recognitions over the years, including the highest awards of IEEE CSS, IFAC, AACC, and ISDG, the IEEE Control Systems Award, and a number of international honorary doctorates and professorships. He was the Editor-in-Chief of Automatica from 2004 to 2014, and is currently an editor of several book series. He is a member of the U.S. National Academy of Engineering and a Fellow of the American Academy of Arts and Sciences, as well as a Fellow of IFAC and SIAM.

Shubham Aggarwal (Graduate Student Member, IEEE) received the B.Tech. and M.Tech. degrees in electrical engineering from the Indian Institute of Technology (BHU) Varanasi, Varanasi, India. He is currently pursuing the Ph.D. degree with the Coordinated Science Laboratory and the Department of Mechanical Science and Engineering, University of Illinois Urbana–Champaign, IL, USA. His current research includes mean-field games and networked control systems.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.