IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Federated TD Learning over Finite-Rate Erasure Channels: Linear Speedup under Markovian Sampling

Nicolò Dal Fabbro, Aritra Mitra and George J. Pappas

***Abstract*— Federated learning (FL) has recently gained much attention due to its effectiveness in speeding up supervised learning tasks under communication and privacy constraints. However, whether similar speedups can be established for reinforcement learning remains much less understood theoretically. Towards this direction, we study a federated policy evaluation problem where agents communicate via a central aggregator to expedite the evaluation of a common policy. To capture typical communication constraints in FL, we consider finite capacity up-link channels that can drop packets based on a Bernoulli erasure model. Given this setting, we propose and analyze `QFedTD` - a quantized federated temporal difference learning algorithm with linear function approximation. Our main technical contribution is to provide a finite-sample analysis of `QFedTD` that (i) highlights the effect of quantization and erasures on the convergence rate; and (ii) establishes a linear speedup w.r.t. the number of agents under Markovian sampling. Notably, while different quantization mechanisms and packet drop models have been extensively studied in the FL, distributed optimization, and networked control systems literature, our work is the first to provide a non-asymptotic analysis of their effects in multi-agent and federated reinforcement learning.**

I. Introduction

Is it possible to obtain statistical models of high accuracy for supervised learning problems (e.g., regression, classification, etc.) by aggregating information from multiple devices while keeping the raw data on these devices private? This is the central question of interest in the popular machine learning paradigm of federated learning (FL) [1]. When the data-generating distributions of the participating devices are identical (or sufficiently similar), several works have shown that one can reap the benefits of collaboration by exchanging locally trained models via a central aggregator (server) [2], [3]. In practice, these models are typically high-dimensional and need to be exchanged over unreliable communication links of limited bandwidth. As such, a large body of work in FL has investigated the effects of uploading quantized models (or model-differentials, i.e., gradients) over channels prone to packet drops/erasures [4], [5]. Drawing inspiration from this literature, in this paper, we ask: *Can we establish collaborative performance gains for federated reinforcement learning (FRL) problems subject to similar communication challenges?* As it turns out, little to nothing is known about this question from a theoretical standpoint.

Towards this direction, we study one of the most basic problems in RL, namely *policy evaluation*, in a federated setting. Specifically, in our problem, N agents, each of whom interacts with the same Markov Decision Process (MDP), communicate via a server to evaluate a fixed policy. While each agent can evaluate the policy on its own using Monte-Carlo sampling or temporal difference (TD) learning algorithms [6], [7], the reason for communicating is the same as in the standard FL setting: *to achieve an N-fold speedup in the sample-complexity of policy evaluation relative to when an agent acts alone*. In the recent survey paper on FRL [8], the authors mention that the goal of the FRL framework is to achieve such speedups while respecting privacy constraints, i.e., without revealing the raw data (states, actions, and rewards) of the agents. Relative to the FL setting, proving finite-time rates for FRL is significantly more challenging since we need to deal with temporally correlated Markovian samples. Indeed, even for the single-agent setting, finite-time rates under Markovian sampling have only recently been established [9]–[12]. For the multi-agent setting, almost all the prior works on TD learning make a restrictive i.i.d. sampling assumption [13], [14]. The only two exceptions to this are the very recent papers [15], [16] that establish linear speedups under Markovian sampling; however, none of the above works consider any communication constraints. As such, establishing linear speedups in FRL under Markovian sampling and communication constraints remains largely unexplored. In this regard, our contributions are as follows.

Contributions. Our first contribution is to formulate a federated policy evaluation problem under two practical constraints on the communication channels: finite capacity and packet drops (lossy links). To capture these constraints, we propose and analyze `QFedTD` - a federated TD algorithm with linear function approximation where agents upload quantized TD update directions to the server over Bernoulli erasure channels [17], [18]. While various quantization and erasure models have been extensively analyzed in the FL [4], distributed optimization [19], and networked control literature [17], [18] for almost two decades, our work is the first to formally study them in multi-agent/federated RL.

Our second and most significant contribution is to provide a rigorous non-asymptotic analysis of `QFedTD` that clearly highlights the effects of quantization and erasures, and establishes an N-fold linear speedup in sample-complexity relative to the single-agent setting. Since RL algorithms often require several samples to achieve acceptable accuracy, our speedup result under realistic communication models is of significant practical importance. We now comment on some

N. Dal Fabbro is with the Department of Information Engineering, University of Padova. Email: dalfabbron@dei.unipd.it. A. Mitra is with the Electrical and Computer Engineering Department, North Carolina State University. Email: amitra2@ncsu.edu. G. J. Pappas is with the Electrical and Systems Engineering Department, University of Pennsylvania. Email: pappasg@seas.upenn.edu. This work was supported by NSF Award 1837253, and the Italian Ministry of Education, University and Research through the PRIN project no. 2017NS9FEY.

of the highlights of our analysis relative to [15] and [16]. Our work crucially departs from both these papers in that, in addition to correlated Markovian samples, we need to contend with two other sources of randomness: one due to randomized quantization and the other due to the Bernoulli packet-dropping processes. Even in the absence of communication challenges, our analysis has the following key benefits. Unlike [16], our work does not employ any projection step. Moreover, compared to the analysis in [15] that relies on Generalized Moreau Envelopes, our proof is significantly shorter and simpler. As a byproduct of this simpler analysis, we derive bounds that have a tighter linear dependence on the mixing time (consistent with the centralized setting) as opposed to the quadratic dependence in [15], [16]; see Section III for more discussion on this point.

II. System Model and Problem Formulation

We consider a setting involving N agents, where all agents interact with the *same* Markov Decision Process (MDP). Let us denote the shared MDP by $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ is a finite state space of size n, $\mathcal{A}$ is a finite action space, $\mathcal{P}$ is a set of action-dependent Markov transition kernels, $\mathcal{R}$ is a reward function, and $\gamma \in (0, 1)$ is the discount factor. We are interested in a *policy evaluation* (PE) problem where the agents exchange information via a central aggregator (server) to evaluate the value function associated with a policy μ. Here, the policy is a map from the states to the actions, i.e., $\mu : \mathcal{S} \to \mathcal{A}$. In what follows, we first briefly review some key concepts relevant to PE with function approximation. Then, we formally describe our communication model, objectives, and technical challenges.

Policy Evaluation with Linear Function Approximation. The policy μ to be evaluated induces a Markov Reward Process (MRP) with transition matrix $\mathbb{P}_\mu$ and reward function $R_\mu : \mathcal{S} \to \mathbb{R}$. The purpose of PE is to evaluate the value function $\boldsymbol{V}_\mu(s)$ for each $s \in \mathcal{S}$, where $\boldsymbol{V}_\mu(s)$ is the discounted expected cumulative reward obtained by playing policy μ starting from initial state s. Formally,

$$\boldsymbol{V}_\mu(s) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k R_\mu(s_k) | s_0 = s\right], \qquad (1)$$

where s_k represents the state of the Markov chain at the discrete time-step k under the action of the policy μ. Our particular interest is in the RL setting where the Markov transition kernels and reward functions are *unknown*.

In several large-scale practical settings, the size n of the state space $\mathcal{S}$ is large, thereby creating a major computational challenge. To work around this issue, we will resort to the popular idea of linear function approximation where $\boldsymbol{V}_\mu$ is approximated by vectors in a linear subspace of $\mathbb{R}^n$ spanned by a set of m basis vectors $\{\boldsymbol{\phi}_\ell\}_{\ell \in [m]}$[1]; importantly, $m \ll n$. To be more precise, let us define the feature matrix $\boldsymbol{\Phi} \triangleq [\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_m] \in \mathbb{R}^{n \times m}$. Given a weight (model) vector $\boldsymbol{\theta} \in \mathbb{R}^m$, the parametric approximation $\hat{\boldsymbol{V}}_{\boldsymbol{\theta}}$ of $\boldsymbol{V}_\mu$ is then given by $\boldsymbol{V}(\boldsymbol{\theta}) := \hat{\boldsymbol{V}}_{\boldsymbol{\theta}} = \boldsymbol{\Phi}\boldsymbol{\theta}$. If we denote the s-th row of $\boldsymbol{\Phi}$ as $\boldsymbol{\phi}'_s$, then the approximation of $\boldsymbol{V}_\mu(s)$, in particular, is given by $\hat{\boldsymbol{V}}_{\boldsymbol{\theta}}(s) = \langle \boldsymbol{\theta}, \boldsymbol{\phi}'_s \rangle$. Throughout, we will make the standard assumption [9] that the columns of $\boldsymbol{\Phi}$ are independent and that the rows are normalized, i.e., $\|\boldsymbol{\phi}'_s\|_2^2 \leq 1, \forall s \in \mathcal{S}$.

Communication Model and `QFedTD` Algorithm. Given the above setup, the goal of the server-agent system is to collectively estimate the model vector $\boldsymbol{\theta}^*$ corresponding to the best linear approximation of $\boldsymbol{V}_\mu$ in the span of $\boldsymbol{\Phi}$. To achieve this goal, we now describe a multi-agent variant of the classical TD(0) algorithm [6]. All agents start out from a common initial state $s_0 \in \mathcal{S}$ with an initial estimate $\boldsymbol{\theta}_0 \in \mathbb{R}^m$. Subsequently, at each time-step $k \in \mathbb{N}$, a global model vector $\boldsymbol{\theta}_k$ is broadcasted by the server to all agents. Each agent $i \in [N]$ then takes an action $a_{i,k} = \mu(s_{i,k})$, and observes the next state $s_{i,k+1} \sim \mathbb{P}_\mu(\cdot|s_{i,k})$ and instantaneous reward $r_{i,k} = R_\mu(s_{i,k})$; here, $s_{i,k}$ is the state of agent i at time-step k. Using the model vector $\boldsymbol{\theta}_k$ and the observation tuple $o_{i,k} = (s_{i,k}, r_{i,k}, s_{i,k+1})$, agent i computes the following local TD update direction:

$$\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k}) = (r_{i,k} + \gamma\langle \boldsymbol{\phi}'_{s_{i,k+1}}, \boldsymbol{\theta}_k\rangle - \langle \boldsymbol{\phi}'_{s_{i,k}}, \boldsymbol{\theta}_k\rangle)\boldsymbol{\phi}'_{s_{i,k}}.$$

We will often use $\mathbf{g}_{i,k}(\boldsymbol{\theta}_k)$ as a shorthand for $\mathbf{g}_{i,k}(\boldsymbol{\theta}_k, o_{i,k})$. Note that although all agents play the same policy μ, and interact with the same MDP, the realizations of the local observation sequences $\{o_{i,k}\}$ can differ across agents. We assume that these observation sequences are *statistically independent* across agents.[2] Intuitively, based on this independence property, one can expect that exchanging agents' local TD update directions should help reduce the variance in the estimate of $\boldsymbol{\theta}^*$. This is precisely where the communication-induced challenges we describe below play a role.

Channel Effects. We model two key aspects of realistic communication channels in large-scale FL settings: finite capacity (due to limited bandwidth) and erasures/packet drops. To account for the first issue, we will employ a simple unbiased quantizer which is a (potentially random) mapping $\mathcal{Q} : \mathbb{R}^m \to \mathbb{R}^m$ satisfying the following constraints [5].

Definition 1. *(Unbiased Quantizer) We say that a quantizer $\mathcal{Q}$ is unbiased if the following hold for all $\mathbf{x} \in \mathbb{R}^m$: (i) $\mathbb{E}[\mathcal{Q}(\mathbf{x})] = \mathbf{x}$, and (ii) there exists some constant $\zeta \geq 0$ such that $\mathbb{E}\left[\|\mathcal{Q}(\mathbf{x}) - \mathbf{x}\|_2^2\right] \leq \zeta\|\mathbf{x}\|_2^2$, where the expectation is w.r.t. the randomness of the quantizer.*

The constant ζ captures the amount of distortion introduced by the quantizer. Using *any* quantizer that satisfies Definition 1, each agent i computes an encoded version $\mathbf{h}_{i,k}(\boldsymbol{\theta}_k) = \mathcal{Q}(\mathbf{g}_{i,k}(\boldsymbol{\theta}_k))$ of $\mathbf{g}_{i,k}(\boldsymbol{\theta}_k)$. Here, we assume that the randomness of the quantizer is independent across agents and also independent of the Markovian observation tuples.

Next, to capture packet drops, we assume that the encoded TD directions are uploaded to the server over Bernoulli erasure channels. Specifically, the transmission of information from an agent i to the server is over a channel whose statistics are governed by an i.i.d. random process $\{b_{i,k}\}$, where for

[1] Given a positive integer m, we use the notation $[m] = 1, ..., m$.

[2] For each agent i, the observations over time are, however, correlated since they are all part of a single Markov chain.

each k, $b_{i,k}$ follows a Bernoulli fading distribution, i.e., $b_{i,k}=0$ with erasure probability $(1-p)$, and $b_{i,k}=1$ with probability p. The packet-dropping processes are assumed to be independent of all other sources of randomness in our model. We are now in a position to describe the global model-update rule at the server:

$$\boxed{\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_k+\alpha\mathbf{v}_k;\;\; \mathbf{v}_k=\frac{1}{N}\sum_{i=1}^{N} b_{i,k}\mathbf{h}_{i,k}(\boldsymbol{\theta}_k),} \tag{2}$$

where α is a constant step-size/learning rate. We refer to the overall updating scheme described above as the Quantized Federated TD learning algorithm, or simply `QFedTD`.

Objective and Challenges. The main goal of this paper is to provide a *finite-time analysis* of `QFedTD`. This is non-trivial for several reasons. Even in the single-agent setting, providing a non-asymptotic analysis of `TD`(0) without any projection step is known to be quite challenging due to temporal correlations between the Markov samples. To analyze `QFedTD`, *we need to contend with three distinct sources of randomness*: (i) randomness due to the temporally correlated Markov samples $\{o_{i,k}\}_{i\in[N]}$; (ii) randomness due to the quantization step; and (iii) randomness due to the Bernoulli packet dropping processes $\{b_{i,k}\}_{i\in[N]}$. Furthermore, unlike a single-agent setting, our goal is to establish a "linear speedup" w.r.t. the number of agents under the different sources of randomness above. This necessitates a very careful analysis that we provide in the subsequent sections.

III. Main Result

In this section, we state and discuss our main result pertaining to the non-asymptotic performance of `QFedTD`. First, however, we need some technical preparation. As is standard, we assume that the rewards are uniformly bounded, i.e., $\exists \bar{r}>0$ such that $R_\mu(s)\le \bar{r}, \forall s\in\mathcal{S}$. This ensures that the value function in (1) is well-defined. Next, we make a standard assumption that plays a key role in the finite-time analysis of TD learning algorithms [7], [9], [10].

Assumption 1. *The Markov chain induced by the policy μ is aperiodic and irreducible.*

Assumption 1 implies that the Markov chain induced by μ admits a unique stationary distribution π [20]. Let $\boldsymbol{\Sigma}=\boldsymbol{\Phi}^\top\mathbf{D}\boldsymbol{\Phi}$, where $\mathbf{D}$ is a diagonal matrix with entries given by the elements of π. Since $\boldsymbol{\Phi}$ is assumed to be full column rank, $\boldsymbol{\Sigma}$ is full rank with a strictly positive smallest eigenvalue $\omega<1$; ω will later show up in our convergence bounds. Next, we define the steady-state local TD direction $\forall\boldsymbol{\theta}\in\mathbb{R}^m$:

$$\bar{\mathbf{g}}(\boldsymbol{\theta})\triangleq\mathbb{E}_{s_{i,k}\sim\pi, s_{i,k+1}\sim\mathbb{P}_\mu(\cdot|s_{i,k})}[\mathbf{g}_{i,k}(\boldsymbol{\theta},o_{i,k})]. \tag{3}$$

Essentially, the *deterministic* recursion $\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_k+\alpha\bar{\mathbf{g}}(\boldsymbol{\theta}_k)$ captures the limiting behavior of the `TD`(0) update rule. In [9], it was shown that the iterates generated by this recursion converge exponentially fast to $\boldsymbol{\theta}^*$, where $\boldsymbol{\theta}^*$ is the unique solution of the projected Bellman equation $\Pi_{\mathbf{D}}\mathcal{T}_\mu(\boldsymbol{\Phi}\boldsymbol{\theta}^*)=\boldsymbol{\Phi}\boldsymbol{\theta}^*$. Here, $\Pi_{\mathbf{D}}(\cdot)$ is the projection operator onto the subspace spanned by $\{\boldsymbol{\phi}_\ell\}_{\ell\in[m]}$ with respect to the inner product $\langle\cdot,\cdot\rangle_{\mathbf{D}}$, and $\mathcal{T}_\mu:\mathbb{R}^n\to\mathbb{R}^n$ is the policy-specific Bellman operator [7]. We now define the notion of mixing time τ_ϵ that will play a crucial role in our analysis.

Definition 2. *Let τ_ϵ be the minimum time step such that $\|\mathbb{E}[\mathbf{g}_{i,k}(\boldsymbol{\theta},o_{i,k})|o_{i,0}]-\bar{\mathbf{g}}(\boldsymbol{\theta})\|\le\epsilon(\|\boldsymbol{\theta}\|+1), \forall k\ge\tau_\epsilon, \forall\boldsymbol{\theta}\in\mathbb{R}^m, \forall i\in[N], \forall o_{i,0}$.*[3]

Assumption 1 implies that the Markov chain induced by μ mixes at a geometric rate [20], i.e., the total variation distance between $\mathbb{P}(s_{i,k}=\cdot|s_{i,0}=s)$ and the stationary distribution π decays exponentially fast $\forall k\ge 0, \forall i\in[N], \forall s\in\mathcal{S}$. This immediately implies the existence of some $K\ge 1$ such that τ_ϵ in Definition 2 satisfies $\tau_\epsilon\le K\log(\frac{1}{\epsilon})$ [11]. Loosely speaking, this means that for a fixed $\boldsymbol{\theta}$, if we want the noisy TD update direction to be ϵ-close (relative to $\boldsymbol{\theta}$) to the steady-state TD direction (where both these directions are evaluated at $\boldsymbol{\theta}$), then the amount of time we need to wait for this to happen scales logarithmically in the precision ϵ. For our purpose, we will set $\epsilon=\alpha^q$, where q is an integer satisfying $q\ge 2$. Unlike the centralized setting where $q=1$ suffices [9], [10], to establish the linear speedup property, we will require $q\ge 2$. Henceforth, we will drop the subscript of $\epsilon=\alpha^q$ in τ_ϵ and simply refer to τ as the mixing time. Let us define by $\sigma\triangleq\max\{1,\bar{r},\|\boldsymbol{\theta}^*\|\}$ the "variance" of the observation model for our problem. Finally, let $\zeta'\triangleq\max\{1,\zeta\}$, where ζ is as in Definition 1, and $\delta_k^2\triangleq\|\boldsymbol{\theta}^*-\boldsymbol{\theta}_k\|^2$. We can now state our main result.

Theorem 1. *Consider the update rule of `QFedTD` in (2). There exist universal constants $C_0, C_2, C_3\ge 1$, such that with $\alpha\le\frac{\omega(1-\gamma)}{C_0\tau\zeta'}$, the following holds for $T\ge 2\tau$:*

$$\mathbb{E}\left[\delta_T^2\right]\le(1-\alpha\omega(1-\gamma)p)^T C_1+\frac{\tau\sigma^2}{\omega(1-\gamma)}\left(\frac{C_2\alpha\zeta'}{N}+C_3\alpha^3\right), \tag{4}$$

where $C_1=4\delta_0^2+2p\sigma^2$.

Discussion: There are several important takeaways from Theorem 1. From (4), we first note that `QFedTD` guarantees linear convergence (in expectation) to a ball around $\boldsymbol{\theta}^*$ whose radius depends on the variance σ^2 of the noise model. While the linear convergence rate gets slackened by the probability of successful transmission p, the "variance term", namely the second term in (4), gets inflated by the quantization parameter ζ. Both of these channel effects are consistent with what one observes for analogous settings in FL [4]. Next, compared to the centralized setting [10, Theorem 7], the variance term in (4) gets scaled down by a factor of N, up to a higher-order $O(\alpha^3)$ term that can be dominated by the (α/N) term for small enough α. Before we make this point explicit, it is worth noting that our variance bound exhibits a tighter dependence on the mixing time τ relative to [15] and [16], where similar bounds are established. In particular, while this dependence is $O(\tau)$ for us, it is $O(\tau^2)$ in [15, Theorem 4.1] and in [16, Theorem 4]. Notably, the $O(\tau)$ dependence that we establish is consistent with results on centralized TD learning [9], [10], and is in fact the optimal

[3] Unless otherwise specified, we use $\|\cdot\|$ to denote the Euclidean norm.

dependence on τ under Markovian data [21]. We now show that with a suitable choice of step-size α, if the number of iterations T is chosen to be large enough, then the mean-square error of `QFedTD` converges exactly to 0 at a rate of $O(1/(NT))$, i.e., we obtain a linear speedup in sample-complexity w.r.t. the number of agents N. To see this, let

$$\alpha = \frac{\log NT}{\omega(1-\gamma)pT}, \quad \text{and} \quad T \geq \frac{2C_0 N\tau\zeta' \log NT}{\omega^2(1-\gamma)^2 p}. \tag{5}$$

We can then easily show that (see our extended technical report [22] for the details)

$$\mathbb{E}[\delta_T^2] \leq O\left(\left(\frac{\zeta'}{p}\right)\frac{\max\{\delta_0^2, \sigma^2\}\tau\log(NT)}{\omega^2(1-\gamma)^2 NT}\right). \tag{6}$$

As far as we are aware, our work is the first to establish a linear speedup result of the above form under Markovian sampling and communication constraints.

IV. Proof of the Main Result

In this section, we prove Theorem 1. We start by introducing the following definitions to lighten the notation:

$$\begin{aligned} \eta^{(i)}_{k,\tau}(\boldsymbol{\theta}) &\triangleq \|\mathbb{E}\left[\mathbf{g}_{i,k}(\boldsymbol{\theta}, o_{i,k})|o_{i,k-\tau}\right] - \bar{\mathbf{g}}(\boldsymbol{\theta})\|, \;\; k \geq \tau, \\ \delta_{k,\tau} &\triangleq \|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-\tau}\|, \;\; k \geq \tau. \end{aligned} \tag{7}$$

For our analysis, we will need the following result from [9].

Lemma 1. *The following holds* $\forall \boldsymbol{\theta} \in \mathbb{R}^m$*:*

$$\langle \boldsymbol{\theta}^* - \boldsymbol{\theta}, \bar{\mathbf{g}}(\boldsymbol{\theta})\rangle \geq \omega(1-\gamma)\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2.$$

We will also use the fact that the random TD update directions and their steady-state versions are 2-Lipschitz [9], i.e., $\forall i \in [N], \forall k \in \mathbb{N}$, and $\forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^m$, we have:

$$\max\{\|\mathbf{g}_{i,k}(\boldsymbol{\theta}) - \mathbf{g}_{i,k}(\boldsymbol{\theta}')\|, \|\bar{\mathbf{g}}(\boldsymbol{\theta}) - \bar{\mathbf{g}}(\boldsymbol{\theta}')\|\} \leq 2\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|. \tag{8}$$

From [10], we also have that $\forall i \in [N], \forall k \in \mathbb{N}, \forall \boldsymbol{\theta} \in \mathbb{R}^m$:

$$\|\mathbf{g}_{i,k}(\boldsymbol{\theta}, o_{i,k})\| \leq 2\|\boldsymbol{\theta}\| + 2\bar{r}. \tag{9}$$

Equipped with the above basic results, we now provide an outline of our proof before delving into the technical details.

Outline of the proof. We start by defining:

$$\begin{aligned} \bar{\mathbf{g}}_N(\boldsymbol{\theta}_k) &\triangleq \frac{1}{N}\sum_{i=1}^{N} b_{i,k}\bar{\mathbf{g}}(\boldsymbol{\theta}_k), \quad \text{and} \\ \psi_k &\triangleq \langle \mathbf{v}_k - \bar{\mathbf{g}}_N(\boldsymbol{\theta}_k), \boldsymbol{\theta}_k - \boldsymbol{\theta}^*\rangle. \end{aligned} \tag{10}$$

Since for all $i \in [N]$, $b_{i,k}$ is independent of $\boldsymbol{\theta}_k$, we have $\mathbb{E}\left[\langle \bar{\mathbf{g}}_N(\boldsymbol{\theta}_k), \boldsymbol{\theta}_k - \boldsymbol{\theta}^*\rangle\right] = p\mathbb{E}\left[\langle \bar{\mathbf{g}}(\boldsymbol{\theta}_k), \boldsymbol{\theta}_k - \boldsymbol{\theta}^*\rangle\right]$. Thus, recalling that $\delta_k^2 \triangleq \|\boldsymbol{\theta}^* - \boldsymbol{\theta}_k\|^2$, and using (2), we obtain

$$\begin{aligned} \mathbb{E}\left[\delta_{k+1}^2\right] &= \mathbb{E}\left[\delta_k^2\right] - 2\alpha\mathbb{E}\left[\langle \boldsymbol{\theta}^* - \boldsymbol{\theta}_k, \mathbf{v}_k\rangle\right] + \alpha^2\mathbb{E}\left[\|\mathbf{v}_k\|^2\right] \\ &= \mathbb{E}\left[\delta_k^2\right] - 2\alpha p\mathbb{E}\left[\langle \boldsymbol{\theta}^* - \boldsymbol{\theta}_k, \bar{\mathbf{g}}(\boldsymbol{\theta}_k)\rangle\right] \\ &\quad + 2\alpha\mathbb{E}\left[\psi_k\right] + \alpha^2\mathbb{E}\left[\|\mathbf{v}_k\|^2\right]. \end{aligned} \tag{11}$$

The main technical burden in proving Theorem 1 is in bounding $\mathbb{E}\left[\|\mathbf{v}_k\|^2\right]$ and $\mathbb{E}\left[\psi_k\right]$ in the above recursion. Following the centralized analysis in [9], [10], one can easily bound $\mathbb{E}\left[\|\mathbf{v}_k\|^2\right]$ using (9). However, this approach will fall short of yielding the desired linear speedup property. Hence, to bound $\mathbb{E}\left[\|\mathbf{v}_k\|^2\right]$, we need a much finer analysis, one that we provide in Lemma 2. Leveraging Lemma 2, we then establish an intermediate result in Lemma 3 that bounds $\mathbb{E}\left[\|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-\tau}\|\right]$. This result, in turn, helps us bound $\mathbb{E}\left[\psi_k\right]$ in Lemma 4. We now proceed to flesh out these steps; some routine calculations are omitted and can be found in [22]. In what follows, $\tau = \tau_\epsilon$ with $\epsilon = \alpha^q$, $q \geq 2$.

Lemma 2. ***(Key Technical Result)*** *For* $k \geq \tau$*, we have*

$$\mathbb{E}\left[\|\mathbf{v}_k\|^2\right] \leq 60\zeta' p\mathbb{E}\left[\delta_k^2\right] + 12\sigma^2 p\left(10\frac{\zeta'}{N} + \alpha^{2q}\right). \tag{12}$$

Proof. Note that $\|\mathbf{v}_k\|^2 \leq \frac{3}{N^2}(T_1 + T_2 + T_3)$, with

$$\begin{aligned} T_1 &= \|\sum_{i=1}^{N} b_{i,k}\mathbf{g}_{i,k}(\boldsymbol{\theta}^*)\|^2, \\ T_2 &= \|\sum_{i=1}^{N} b_{i,k}(\mathbf{g}_{i,k}(\boldsymbol{\theta}_k) - \mathbf{g}_{i,k}(\boldsymbol{\theta}^*))\|^2, \quad \text{and} \\ T_3 &= \|\sum_{i=1}^{N} b_{i,k}(\mathbf{g}_{i,k}(\boldsymbol{\theta}_k) - \mathbf{h}_{i,k}(\boldsymbol{\theta}_k))\|^2. \end{aligned} \tag{13}$$

We now proceed to bound $T_1 - T_3$. To that end, we first write T_1 as $T_1 = T_{11} + T_{12}$, with $T_{11} = \sum_{i=1}^{N} b_{i,k}^2\|\mathbf{g}_{i,k}(\boldsymbol{\theta}^*)\|^2$, and $T_{12} = \sum_{i,j=1, i\neq j}^{N} b_{i,k}b_{j,k}\langle \mathbf{g}_{i,k}(\boldsymbol{\theta}^*), \mathbf{g}_{j,k}(\boldsymbol{\theta}^*)\rangle$. Now using (9), we obtain $T_{11} \leq 8(\|\boldsymbol{\theta}^*\|^2 + \bar{r}^2)\sum_{i=1}^{N} b_{i,k}^2$. Recalling that $\sigma \triangleq \max\{1, \bar{r}, \|\boldsymbol{\theta}^*\|\}$, we then have $\mathbb{E}\left[T_{11}\right] \leq 16\sigma^2\mathbb{E}\left[\sum_{i=1}^{N} b_{i,k}^2\right] = 16\sigma^2 Np$. Next, to bound the cross-terms in T_{12}, we will exploit the mixing property in Definition 2. To that end, we note that since (i) $\bar{\mathbf{g}}(\boldsymbol{\theta}^*) = \mathbf{0}$ [9], (ii) the packet-dropping processes are independent of the Markovian tuples, and (iii) $\mathbf{g}_{i,k}(\boldsymbol{\theta}^*)$ and $\mathbf{g}_{j,k}(\boldsymbol{\theta}^*)$ are independent for $i \neq j$,

$$\begin{aligned} \mathbb{E}\left[T_{12}\right] = \sum_{\substack{i,j=1 \\ i\neq j}}^{N} \mathbb{E}\left[b_{i,k}b_{j,k}\right]\langle \mathbb{E}\left[\mathbb{E}\left[\mathbf{g}_{i,k}(\boldsymbol{\theta}^*)|o_{i,k-\tau}\right] - \bar{\mathbf{g}}(\boldsymbol{\theta}^*)\right], \\ \mathbb{E}\left[\mathbb{E}\left[\mathbf{g}_{j,k}(\boldsymbol{\theta}^*)|o_{j,k-\tau}\right] - \bar{\mathbf{g}}(\boldsymbol{\theta}^*)\right]\rangle. \end{aligned}$$

Using the Cauchy-Schwarz inequality followed by Jensen's inequality, we can further bound the above inner-product via $\mathbb{E}\left[\eta^{(i)}_{k,\tau}(\boldsymbol{\theta}^*)\right] \times \mathbb{E}\left[\eta^{(j)}_{k,\tau}(\boldsymbol{\theta}^*)\right] \leq 4\sigma^2\alpha^{2q}$. For the last inequality, we used the mixing property by noting that $k \geq \tau$. Combining this analysis with the fact that $\mathbb{E}\left[b_{i,k}b_{j,k}\right] = \mathbb{E}\left[b_{i,k}\right]\mathbb{E}\left[b_{j,k}\right] = p^2$, we obtain that $\mathbb{E}\left[T_{12}\right] \leq 4N^2p^2\sigma^2\alpha^{2q}$. Combining the bounds for $\mathbb{E}\left[T_{11}\right]$ and $\mathbb{E}\left[T_{12}\right]$ thus yields:

$$\mathbb{E}\left[T_1\right] \leq 16\sigma^2 Np + 4N^2p^2\sigma^2\alpha^{2q}. \tag{14}$$

Now, using (8), we see that

$$\begin{aligned} \mathbb{E}\left[T_2\right] &\leq N\sum_{i=1}^{N}\mathbb{E}\left[b_{i,k}^2\|\mathbf{g}_{i,k}(\boldsymbol{\theta}_k) - \mathbf{g}_{i,k}(\boldsymbol{\theta}^*)\|^2\right] \\ &\leq 4N\mathbb{E}\left[\delta_k^2\right]\sum_{i=1}^{N}\mathbb{E}\left[b_{i,k}^2\right] = 4pN^2\mathbb{E}\left[\delta_k^2\right]. \end{aligned} \tag{15}$$

Defining $\boldsymbol{\lambda}_{i,k}(\boldsymbol{\theta}_k) \triangleq \mathbf{h}_{i,k}(\boldsymbol{\theta}_k) - \mathbf{g}_{i,k}(\boldsymbol{\theta}_k)$, we now turn to bounding T_3 by writing it as $T_3 = T_{31} + T_{32}$ where,

$$\begin{aligned} T_{31} &= \sum_{i=1}^{N} b_{i,k}^2 \|\boldsymbol{\lambda}_{i,k}(\boldsymbol{\theta}_k)\|^2, \quad \text{and} \\ T_{32} &= \sum_{\substack{i,j \\ i \neq j}}^{N} b_{i,k} b_{j,k} \langle \boldsymbol{\lambda}_{i,k}(\boldsymbol{\theta}_k), \boldsymbol{\lambda}_{j,k}(\boldsymbol{\theta}_k) \rangle. \end{aligned} \tag{16}$$

We now proceed to bound $\mathbb{E}[T_{31}]$ and $\mathbb{E}[T_{32}]$ as follows:

$$\begin{aligned} \mathbb{E}[T_{31}] &= \sum_{i=1}^{N} \mathbb{E}\left[b_{i,k}^2\right] \mathbb{E}\left[\mathbb{E}\left[\|\boldsymbol{\lambda}_{i,k}(\boldsymbol{\theta}_k)\|^2 | o_{i,k}, \boldsymbol{\theta}_k\right]\right] \\ &\overset{(a)}{\leq} \sum_{i=1}^{N} p\zeta \mathbb{E}\left[\|\mathbf{g}_{i,k}(\boldsymbol{\theta}_k)\|^2\right] \\ &\overset{(b)}{\leq} 8Np\zeta(\mathbb{E}\left[\|\boldsymbol{\theta}_k\|^2\right] + \sigma^2) \\ &\leq 16Np\zeta \mathbb{E}\left[\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\|^2\right] + 24Np\zeta\sigma^2, \end{aligned}$$

where (a) follows from the variance bound of the quantizer map $\mathcal{Q}(\cdot)$, and (b) follows from (9). Next, observe that:

$$\mathbb{E}[T_{32}] = p^2 \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mathbb{E}\left[\mathbb{E}\left[\langle \boldsymbol{\lambda}_{i,k}(\boldsymbol{\theta}_k), \boldsymbol{\lambda}_{j,k}(\boldsymbol{\theta}_k) \rangle | o_{i,k}, o_{j,k}, \boldsymbol{\theta}_k\right]\right].$$

Using the fact that the randomness of the quantization map is independent across agents, and the unbiasedness of $\mathcal{Q}(\cdot)$, we conclude that $\mathbb{E}[T_{32}] = 0$. Combining the bounds on $\mathbb{E}[T_1]$, $\mathbb{E}[T_2]$, and $\mathbb{E}[T_3]$ above yields the desired result. □

Later in the analysis, we will once again need to invoke a mixing time argument by conditioning on $\boldsymbol{\theta}_{k-\tau}$. This will give rise to the $\delta_{k,\tau} = \|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-\tau}\|$ term that we proceed to bound below by leveraging Lemma 2.

Lemma 3. *Let $\alpha \leq \frac{1}{484\tau\zeta'}$ and $k \geq 2\tau$. Then, we have*

$$\mathbb{E}\left[\delta_{k,\tau}^2\right] \leq 480\alpha^2\tau^2 p\zeta' \mathbb{E}\left[\delta_k^2\right] + \alpha^2\tau^2 p\sigma^2 \left(\frac{360\zeta'}{N} + 4\alpha^q\right).$$

Proof. We start with a bound on δ_{k+1}^2:

$$\begin{aligned} \delta_{k+1}^2 &= \delta_k^2 - 2\alpha\langle \mathbf{v}_k, \boldsymbol{\theta}^* - \boldsymbol{\theta}_k \rangle + \alpha^2\|\mathbf{v}_k\|^2 \\ &\leq (1+\alpha)\delta_k^2 + (\alpha + \alpha^2)\|\mathbf{v}_k\|^2 \\ &\leq (1+\alpha)\delta_k^2 + 2\alpha\|\mathbf{v}_k\|^2. \end{aligned} \tag{17}$$

Using Lemma 2 and the fact that $p < 1$, we obtain

$$\mathbb{E}\left[\delta_{k+1}^2\right] \leq (1 + 121\alpha\zeta')\mathbb{E}\left[\delta_k^2\right] + \underbrace{24\alpha p\sigma^2 \left(\frac{10\zeta'}{N} + \alpha^{2q}\right)}_{B}.$$

Iterating this inequality, we get for any $k - \tau \leq k' \leq k$,

$$\mathbb{E}\left[\delta_{k'}^2\right] \leq (1 + 121\alpha\zeta')^\tau \mathbb{E}\left[\delta_{k-\tau}^2\right] + B \sum_{\ell=0}^{\tau-1} (1 + 121\alpha\zeta')^\ell. \tag{18}$$

Now using $(1 + x) \leq e^x, \forall x \in \mathbb{R}$, observe that $(1 + 121\alpha\zeta')^\ell \leq (1+121\alpha\zeta')^\tau \leq e^{0.25} \leq 2$, for $\alpha \leq 1/(484\tau\zeta')$. Thus, $\sum_{\ell=0}^{\tau-1}(1+121\alpha\zeta')^\ell \leq 2\tau$. Plugging this bound in (18),

$$\mathbb{E}\left[\delta_{k'}^2\right] \leq 2\mathbb{E}\left[\delta_{k-\tau}^2\right] + 2\tau B. \tag{19}$$

Next, observe that $\delta_{k,\tau}^2 \leq \tau \sum_{\ell=k-\tau}^{k-1} \|\boldsymbol{\theta}_{\ell+1} - \boldsymbol{\theta}_\ell\|^2 = \tau\alpha^2 \sum_{\ell=k-\tau}^{k-1} \|\mathbf{v}_\ell\|^2$. Since $k \geq 2\tau$, we have $\ell \geq \tau$. Hence, we can invoke Lemma 2 to bound $\mathbb{E}\left[\|\mathbf{v}_\ell\|^2\right]$. This yields

$$\mathbb{E}\left[\delta_{k,\tau}^2\right] \leq \alpha^2\tau \sum_{\ell=k-\tau}^{k-1} 60\zeta' p \mathbb{E}\left[\delta_\ell^2\right] + 0.5\alpha\tau^2 B. \tag{20}$$

Using (19) to bound $\mathbb{E}\left[\delta_\ell^2\right]$ above, we further obtain

$$\mathbb{E}\left[\delta_{k,\tau}^2\right] \leq \alpha^2\tau \sum_{\ell=k-\tau}^{k-1} 120\zeta' p \left(\mathbb{E}\left[\delta_{k-\tau}^2\right] + \tau B\right) + \frac{1}{2}\alpha\tau^2 B.$$

Simplifying using $\alpha \leq 1/484\zeta'\tau$, $p < 1$, and $q \geq 2$ yields

$$\mathbb{E}\left[\delta_{k,\tau}^2\right] \leq 120\alpha^2\tau^2 p\zeta' \mathbb{E}\left[\delta_{k-\tau}^2\right] + \alpha^2\tau^2\sigma^2 p \left(\frac{180\zeta'}{N} + 2\alpha^q\right).$$

Using $\delta_{k-\tau}^2 \leq 2\delta_k^2 + 2\delta_{k,\tau}^2$ and $240\alpha^2\tau^2\zeta' \leq 1/2$ to simplify the above inequality, we arrive at the desired result. □

Our next result provides a bound on $\mathbb{E}[\psi_k]$, and is the final ingredient we need to prove Theorem 1.

Lemma 4. *Define $\mathbf{g}_N(\boldsymbol{\theta}_k) \triangleq \frac{1}{N}\sum_{i=1}^{N} b_{i,k}\mathbf{g}_{i,k}(\boldsymbol{\theta}_k)$, and let $\alpha \leq 1/(484\zeta'\tau)$ and $k \geq 2\tau$. We have*

$$\mathbb{E}[\psi_k] \leq \alpha\tau p \left(3191\zeta' \mathbb{E}\left[\delta_k^2\right] + \sigma^2 \left(\frac{2461\zeta'}{N} + 30\alpha^q\right)\right).$$

Proof. We can write $\psi_k = T_1 + T_2 + T_3 + T_4 + T_5$, with

$$\begin{aligned} T_1 &= \langle \boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-\tau}, \mathbf{g}_N(\boldsymbol{\theta}_k) - \bar{\mathbf{g}}_N(\boldsymbol{\theta}_k) \rangle, \\ T_2 &= \langle \boldsymbol{\theta}_{k-\tau} - \boldsymbol{\theta}^*, \mathbf{g}_N(\boldsymbol{\theta}_{k-\tau}) - \bar{\mathbf{g}}_N(\boldsymbol{\theta}_{k-\tau}) \rangle, \\ T_3 &= \langle \boldsymbol{\theta}_{k-\tau} - \boldsymbol{\theta}^*, \mathbf{g}_N(\boldsymbol{\theta}_k) - \mathbf{g}_N(\boldsymbol{\theta}_{k-\tau}) \rangle, \\ T_4 &= \langle \boldsymbol{\theta}_{k-\tau} - \boldsymbol{\theta}^*, \bar{\mathbf{g}}_N(\boldsymbol{\theta}_{k-\tau}) - \bar{\mathbf{g}}_N(\boldsymbol{\theta}_k) \rangle, \\ T_5 &= \langle \boldsymbol{\theta}_k - \boldsymbol{\theta}^*, \mathbf{v}_k - \mathbf{g}_N(\boldsymbol{\theta}_k) \rangle. \end{aligned} \tag{21}$$

Note that $T_1 \leq \frac{1}{2\alpha\tau}\delta_{k,\tau}^2 + \frac{1}{2}\alpha\tau\|\mathbf{g}_N(\boldsymbol{\theta}_k) - \bar{\mathbf{g}}_N(\boldsymbol{\theta}_k)\|^2$, and so

$$T_1 \leq \frac{1}{2\alpha\tau}\delta_{k,\tau}^2 + \alpha\tau\|\mathbf{g}_N(\boldsymbol{\theta}_k)\|^2 + \alpha\tau\|\bar{\mathbf{g}}_N(\boldsymbol{\theta}_k) - \bar{\mathbf{g}}_N(\boldsymbol{\theta}^*)\|^2.$$

Using (8), note that $\|\bar{\mathbf{g}}_N(\boldsymbol{\theta}_k) - \bar{\mathbf{g}}_N(\boldsymbol{\theta}^*)\|^2 \leq \frac{4}{N}\sum_{i=1}^{N} b_{i,k}^2\delta_k^2$. Also, $\|\mathbf{g}_N(\boldsymbol{\theta}_k)\|^2$ can be bounded exactly in the same way as the first two terms in (13) of Lemma 2. Using these bounds and invoking Lemma 3 yields:

$$\mathbb{E}[T_1] \leq 304\alpha\tau\zeta' p \mathbb{E}\left[\delta_k^2\right] + \alpha\tau p\sigma^2 \left(\frac{300\zeta'}{N} + 3\alpha^q\right).$$

Next we bound $\mathbb{E}[T_3]$ and $\mathbb{E}[T_4]$. Observe that:

$$\begin{aligned} \mathbb{E}[T_3] &= \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[b_{i,k}\langle \boldsymbol{\theta}_{k-\tau} - \boldsymbol{\theta}^*, (\mathbf{g}_{i,k}(\boldsymbol{\theta}_k) - \mathbf{g}_{i,k}(\boldsymbol{\theta}_{k-\tau})) \rangle\right] \\ &\leq p\mathbb{E}\left[\delta_{k-\tau}\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{g}_{i,k}(\boldsymbol{\theta}_k) - \mathbf{g}_{i,k}(\boldsymbol{\theta}_{k-\tau})\|\right] \\ &\overset{(8)}{\leq} \frac{\alpha\tau p}{2}\mathbb{E}\left[\delta_{k-\tau}^2\right] + \frac{2p}{\alpha\tau}\mathbb{E}\left[\delta_{k,\tau}^2\right]. \end{aligned}$$

Using $\delta_{k-\tau}^2 \le 2\delta_k^2 + 2\delta_{k,\tau}^2$ and Lemma 3, we then obtain:

$$\mathbb{E}\left[T_3\right] \le 1441\alpha\tau p\zeta'\mathbb{E}\left[\delta_k^2\right] + 6\alpha\tau p\sigma^2\left(\frac{180\zeta'}{N} + 2\alpha^q\right).$$

Using the same process, we can derive the exact same bound for $\mathbb{E}\left[T_4\right]$. We now bound $\mathbb{E}\left[T_2\right]$. For ease of notation, let us define $\mathcal{F}_{k,\tau} = (\{o_{i,k-\tau}\}_{i=1}^N, \boldsymbol{\theta}_{k-\tau})$. Observe:

$$\begin{aligned}\mathbb{E}\left[T_2\right] &= \mathbb{E}\left[\mathbb{E}\left[T_2|\mathcal{F}_{k,\tau}\right]\right] = \mathbb{E}[\langle\boldsymbol{\theta}_{k-\tau} - \boldsymbol{\theta}^*,\\ &\quad \frac{p}{N}\sum_{i=1}^N(\mathbb{E}\left[\mathbf{g}_{i,k}(\boldsymbol{\theta}_{k-\tau}, o_{i,k})|\mathcal{F}_{k,\tau}\right] - \bar{\mathbf{g}}(\boldsymbol{\theta}_{k-\tau}))\rangle]\\ &\le \mathbb{E}\left[\delta_{k-\tau}\frac{p}{N}\sum_{i=1}^N \eta_{k,\tau}^{(i)}(\boldsymbol{\theta}_{k-\tau})\right]\\ &\le p\alpha^q\mathbb{E}\left[\delta_{k-\tau}(1 + \|\boldsymbol{\theta}_{k-\tau}\|)\right].\end{aligned}$$

Since $\alpha < 1$, we have $\delta_{k-\tau}(\delta_{k-\tau} + 2\sigma) \le \frac{\delta_{k-\tau}^2}{\alpha} + 2\sigma\delta_{k-\tau} + \alpha\sigma^2 = \left(\frac{\delta_{k-\tau}}{\sqrt{\alpha}} + \sqrt{\alpha}\sigma\right)^2 \le 2\left(\frac{\delta_{k-\tau}^2}{\alpha} + \alpha\sigma^2\right)$. Using $q \ge 2$,

$$\begin{aligned}\mathbb{E}\left[T_2\right] &\le 2p\alpha^q\mathbb{E}\left[\frac{1}{\alpha}\delta_{k-\tau}^2 + \alpha\sigma^2\right]\\ &\le 2p\alpha\mathbb{E}\left[\delta_{k-\tau}^2\right] + 2p\alpha^{q+1}\sigma^2.\end{aligned} \tag{22}$$

Using $\delta_{k-\tau}^2 \le 2\delta_k^2 + 2\delta_{k,\tau}^2$, Lemma 3, and simplifying yields:

$$\mathbb{E}\left[T_2\right] \le 5\alpha\tau p\zeta'\mathbb{E}\left[\delta_k^2\right] + \alpha\tau p\sigma^2\left(\frac{\zeta'}{N} + 3\alpha^q\right). \tag{23}$$

Finally, to bound T_5, let $\mathcal{F}_k = \{\{o_{i,k}\}_{i=1}^N, \boldsymbol{\theta}_k\}$. We have

$$\mathbb{E}\left[T_5\right] = \mathbb{E}\left[\langle\boldsymbol{\theta}_k - \boldsymbol{\theta}^*, \underbrace{\mathbb{E}\left[\mathbf{v}_k - \mathbf{g}_N(\boldsymbol{\theta}_k)|\mathcal{F}_k\right]}_{T_{51}}\rangle\right]. \tag{24}$$

Note that $T_{51} = \frac{p}{N}\sum_{i=1}^N \mathbb{E}\left[\mathbf{h}_{i,k}(\boldsymbol{\theta}_k) - \mathbf{g}_{i,k}(\boldsymbol{\theta}_k)|\mathcal{F}_k\right] = 0$, based on the unbiasedness of $\mathcal{Q}(\cdot)$. Thus, $\mathbb{E}\left[T_5\right] = 0$. Collecting the bounds on $T_1 - T_5$ concludes the proof. □

With the help of the auxiliary lemmas provided above, we are now ready to prove our main result, i.e., Theorem 1.

Proof of Theorem 1. Letting $\alpha \le \frac{1}{484\zeta'\tau}$, we can apply the bounds in Lemmas 1, 2, and 4 to (11). This yields:

$$\begin{aligned}\mathbb{E}\left[\delta_{k+1}^2\right] &\le \mathbb{E}\left[\delta_k^2\right] - \alpha p(2(1-\gamma)\omega - 6446\alpha\tau\zeta')\mathbb{E}\left[\delta_k^2\right]\\ &\quad + 5162\frac{\alpha^2\tau p\sigma^2\zeta'}{N} + 61\alpha^{(2+q)}\tau p\sigma^2.\end{aligned} \tag{25}$$

For $\alpha \le \frac{\omega(1-\gamma)}{C_0\tau\zeta'}$ with $C_0 = 6446$, we then obtain:

$$\begin{aligned}\mathbb{E}\left[\delta_{k+1}^2\right] &\le (1 - \alpha\omega(1-\gamma)p)\mathbb{E}\left[\delta_k^2\right]\\ &\quad + 5162\frac{\alpha^2\tau p\sigma^2\zeta}{N} + 61\alpha^{(2+q)}\tau p\sigma^2.\end{aligned} \tag{26}$$

Iterating the last inequality, we have $\forall k \ge 2\tau$:

$$\mathbb{E}\left[\delta_k^2\right] \le \rho^{k-2\tau}\mathbb{E}\left[\delta_{2\tau}^2\right] + \frac{\tau\sigma^2}{\omega(1-\gamma)}\left(\frac{C_2\alpha\zeta'}{N} + C_3\alpha^3\right),$$

where $\rho = (1 - \alpha\omega(1-\gamma)p)$, $C_2 = 5162$, $C_3 = 61$, and we set $q = 2$. It only remains to show that with our choice of α, $\mathbb{E}\left[\delta_{2\tau}^2\right] = O(\delta_0^2 + \sigma^2)$. This follows from some simple algebra and steps similar to those in the proof of Lemma 3. We omit these details here; they can be found in [22].

References

[1] J. Konečnỳ, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *arXiv preprint arXiv:1610.02527*, 2016.

[2] S. U. Stich, "Local SGD converges fast and communicates little," *arXiv preprint arXiv:1805.09767*, 2018.

[3] A. Khaled, K. Mishchenko, and P. Richtárik, "Tighter theory for local SGD on identical and heterogeneous data," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 4519–4529.

[4] A. Reisizadeh, A. Mokhtari, H. Hassani, A. Jadbabaie, and R. Pedarsani, "Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization," in *AISTATS*. PMLR, 2020, pp. 2021–2031.

[5] A. Beznosikov, S. Horváth, P. Richtárik, and M. Safaryan, "On biased compression for distributed learning," *arXiv:2002.12410*, 2020.

[6] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine learning*, vol. 3, no. 1, pp. 9–44, 1988.

[7] J. N. Tsitsiklis and B. Van Roy, "An analysis of temporal-difference learning with function approximation," in *IEEE Transactions on Automatic Control*, 1997.

[8] J. Qi, Q. Zhou, L. Lei, and K. Zheng, "Federated reinforcement learning: techniques, applications, and open challenges," *arXiv preprint arXiv:2108.11887*, 2021.

[9] J. Bhandari, D. Russo, and R. Singal, "A finite time analysis of temporal difference learning with linear function approximation," in *Conference on learning theory*. PMLR, 2018, pp. 1691–1692.

[10] R. Srikant and L. Ying, "Finite-time error bounds for linear stochastic approximation and TD learning," in *Conference on Learning Theory*. PMLR, 2019, pp. 2803–2830.

[11] Z. Chen, S. Zhang, T. T. Doan, S. T. Maguluri, and J.-P. Clarke, "Performance of q-learning with linear function approximation: Stability and finite-time analysis," *arXiv preprint arXiv:1905.11425*, p. 4, 2019.

[12] G. Patil, L. Prashanth, D. Nagaraj, and D. Precup, "Finite time analysis of temporal difference learning with linear function approximation: Tail averaging and regularisation," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2023, pp. 5438–5448.

[13] T. Doan, S. Maguluri, and J. Romberg, "Finite-time analysis of distributed TD (0) with linear function approximation on multi-agent reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2019, pp. 1626–1635.

[14] R. Liu and A. Olshevsky, "Distributed TD (0) with almost no communication," *arXiv preprint arXiv:2104.07855*, 2021.

[15] S. Khodadadian, P. Sharma, G. Joshi, and S. T. Maguluri, "Federated reinforcement learning: Linear speedup under markovian sampling," in *ICML*. PMLR, 2022, pp. 10 997–11 057.

[16] H. Wang, A. Mitra, H. Hassani, G. J. Pappas, and J. Anderson, "Federated temporal difference learning with linear function approximation under environmental heterogeneity," *arXiv:2302.02212*, 2023.

[17] C. N. Hadjicostis and R. Touri, "Feedback control utilizing packet dropping network links," in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, vol. 2. IEEE, 2002, pp. 1205–1210.

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

[19] M. G. Rabbat and R. D. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 798–808, 2005.

[20] D. A. Levin and Y. Peres, *Markov chains and mixing times*. American Mathematical Soc., 2017, vol. 107.

[21] D. Nagaraj, X. Wu, G. Bresler, P. Jain, and P. Netrapalli, "Least squares regression with markovian data: Fundamental limits and algorithms," *Advances in neural information processing systems*, vol. 33, pp. 16 666–16 676, 2020.

[22] N. Dal Fabbro, A. Mitra, and G. J. Pappas, "Federated TD Learning over Finite-Rate Erasure Channels: Linear Speedup under Markovian Sampling," *arXiv preprint arXiv:2305.08104*, 2023.