Truthful Learning from Biased Agents

Shugang Hao
Singapore University of Technology and Design
Singapore, Singapore
shugang_hao@sutd.edu.sg

Lingjie Duan
Singapore University of Technology and Design
Singapore, Singapore
lingjie_duan@sutd.edu.sg

ABSTRACT

Today it is common for a central system (e.g., headquarter) to request local model updates from multiple agents (branches) to train and deploy a global model in a fair weighted-average way. Yet, selfish agents are found to be biased with their own learning objectives and may strategically misreport to mislead the system's decision to their favors. Without access to agents' local datasets, it is difficult for the system to verify and correct such misreports to learn the grand truth. To our best knowledge, we are the first to study how to truthfully learn from biased agents for training a fair global model for all agents. We find the current practice of the "babbling equilibrium" performs poorly, where the system simply removes unverified agents' updates and replaces with historical belief. Though the popular "weighted median" scheme in the algorithmic game theory literature ensures agents' truthful reporting, we prove that it may perform even worse than the babbling equilibrium. Alternatively, we design a novel PArtial Information Disclosure (PAID) mechanism to truthfully learn from biased agents while minimizing the global learning error. We prove that PAID's learning error is at most 1/4 of the babbling equilibrium and 2/7 of the weighted median scheme, respectively. We also show that our PAID is robust to erroneous agent belief held by the system to yield bounded loss. Finally, we run experiments using a real-world dataset to demonstrate our PAID's great advantages over the other approaches and show its positive "side-effect" to even benefit most biased agents.

CCS CONCEPTS

• **Networks → Network economics**; • **Theory of computation → Multi-agent learning**; **Algorithmic mechanism design**.

KEYWORDS

Truthful learning, Biased agents, Mechanism design.

ACM Reference Format:
Shugang Hao and Lingjie Duan. 2024. Truthful Learning from Biased Agents. In . ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Nowadays, it is common for a central system (e.g., headquarter) to aggregate local model updates from multiple agents (e.g., branches) to train and deploy a global model in a fair weighted-average way (e.g., [2]). Yet, selfish agents are found biased to optimize for their own learning objectives and they may strategically misreport their updates to mislead the system's global deployment decision toward their local models (e.g., [17, 23]). For instance, in constructing a global fraud detection system (e.g., [20]), a branch in a luxury area can downplay the significance of customer income in its model update, thereby skewing the multi-national bank's global decision-making in its favor. Similarly, in building a global survival prediction model on breast cancer patients (e.g., [26]), a clinic expertizing more in early-stage cancer detection can adjust its model update to overemphasize features related to early detection, aiming to mislead the cancer center's global decision toward its local model. Biased agents' misreports pose a significant risk to the global model's correctness and fairness and should be carefully addressed.

In the federated learning (FL) literature, some studies focus on mitigating bias caused by agents' non-IID datasets (e.g., [27]). Abay *et al.* [1] propose a pre-processing bias mitigation method to compute weights of data in the training dataset. Xu *et al.* [29] propose a FL framework that learns to deploy bias-eliminating augmenters for producing client-specific bias-conflicting samples. Tang *et al.* [25] propose an unbiased algorithm by incorporating stratified samplings in FL, which samples clients based on their data statistics to represent the entire population. These studies assume that agents incur the same learning objective as the system and will truthfully upload their model updates. Yet, in practice, biased agents only care about their own learning objectives rather than the entire system. They may strategically misreport their updates to mislead the system's global decision to favor their local situation (e.g., [17, 23].)

Besides, some FL and distributed learning studies aim to develop defense schemes against Byzantine attacks, where attackers send arbitrary model updates to the system (e.g., [6]). Yin *et al.* [31] propose element-wise median and marginal trimmed operations on the gradient updates from the clients to address Byzantine behaviors. Prakash *et al.* [24] propose a method to compute a guiding gradient in every iteration for each client and utilize a per client criteria for filtering out Byzantine updates. Yang *et al.* [30] present a novel decentralized learning method by utilizing the distributed stochastic gradient descent algorithm combined with a novel robust aggregator to address Byzantine attacks, achieving superior performance than existing algorithms. These works assume that Byzantine attackers are irrational and not self-interested, aiming to use arbitrary updates to maximally poison the system's global decision. Differently, here we consider biased agents who aim to selfishly mislead the system into deploying their preferred local models for benefits. Their model updates (though manipulated) may still reveal useful information.

There are two challenges on learning from biased agents for a fair global model. On the system side, without access to agents'

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

local datasets, the system finds it difficult to verify and correct their misreports to learn the grand truth (e.g., [18]). On the agent side, they can strategically manipulate their updates against the system's learning or inference, which makes it even harder to learn the truth. The current practice of the "babbling equilibrium" (e.g., [11]) simply removes unverified agents' updates and replaces with the system's historical belief or inference. Our first question naturally arises:

- *Q1. How bad is the current practice of babbling equilibrium without learning from biased agents? How do the existing truthful mechanisms perform?*

Later we show that the current practice of babbling equilibrium performs poorly. This motivates us to find some other truthful mechanisms. In practice, it may be difficult for the system to implement pricing mechanisms on agents as incentives for data/model sharing systems (e.g., [3, 8]). Thus, it is desirable to design non-monetary truthful information disclosure mechanisms. There are mainly two streams of work on non-monetary mechanisms in the literature. On one hand, there are studies on truthful mechanism design for cheap-talk games, where information senders (i.e., agents) observe the nature state and send messages to proactively affect the receiver's (system's) inference of the actual state (e.g., [4, 14]). Yet, these mechanisms cannot fit into our learning setting. They consider a single agent and assume uniformity in their data, which cannot be used for our scenario with an arbitrary agent number and diverse agents' data or loss objectives. On the other hand, in the algorithmic game theory literature, there are relevant studies on facility location games (e.g., [7, 28]), where the system aims to incentivize customers' truthful reporting of their locations to optimize facility placement and each customer can strategically misreport his location to mislead the facility placement as close to his location. The popular "weighted median" scheme is widely used to return customers' truthful reporting. Yet later we analyze and show that this scheme can perform even worse than the "babbling equilibrium". As such, our second question is:

- *Q2. How to design the very best truthful mechanisms to guide information disclosure from biased agents and achieve the best possible system performance?*

We summarize our key novelty and main results below.

- *Novel truthful learning study from biased agents:* To our best knowledge, we are the first to study how to truthfully learn from biased agents for training a fair global model for all agents. Unlike the FL literature (e.g., [1, 25, 27, 29]), we practically consider that an agent is biased with his own loss function and strategically misreports to mislead the system's decision in his favor. We find the current practice of the "babbling equilibrium" performs poorly, where the system simply removes unverified agents' updates and replaces with historical belief. We aim to present new analytical studies to guide *how a system can best learn from biased agents for the goal of deploying a fair global model.*
- *A novel partial information disclosure mechanism:* Though the popular weighted median scheme ensures agents' truthful reporting, we prove that it may perform even worse than the babbling equilibrium. Alternatively, we design a novel PArtial Information Disclosure (PAID) mechanism to truthfully learn from biased agents while minimizing the global learning error. We transform the PAID design into solving a linear, autonomous, second-order difference equation in closed form for agents' uniformly distributed model parameters. We prove that PAID's error is at most 1/4 of the babbling equilibrium and 2/7 of the weighted median scheme.
- *Robustness to agents' general distributions and the system's erroneous agent belief:* To let our PAID mechanism fit into any distribution of agents' model parameters, we propose an efficient algorithm, which is scalable for a large group of agents and only incurs a quadratic convergence rate. Furthermore, we prove that our PAID is robust to erroneous agent belief held by the system to yield limited loss. Finally, we run experiments using a real-world dataset to demonstrate our PAID's great advantages over the others and show its positive "side-effect" to even benefit most biased agents.

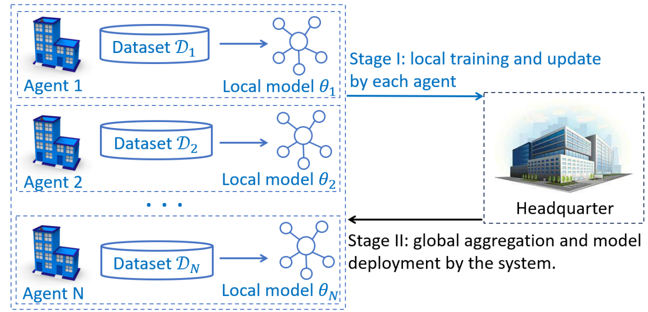

Figure 1: System model illustration in two stages. In Stage I, N agents train their local model parameters $\{\theta_i\}_{i=1}^N$ and message $\{m_i(\theta_i)\}_{i=1}^N$ to the system, where $m_i(\theta_i)$ may be manipulated to differ from θ_i. In Stage II, the system determines the global model $\hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)$ based on the updates to best infer the fair weighted model $\sum_{i=1}^N w_i\theta_i$.

The rest of this paper is organized as follows. Section 2 introduces the system model and the dynamic Bayesian game formulation. Section 3 analyzes two common schemes used in the literature as benchmarks for our PAID to compare later. Section 4 details our PAID design and analysis. Section 5 compares the performance of our PAID with two benchmark schemes and proposes an efficient alogrithm for our PAID design for agents' general distributions. Section 6 extends our PAID design for erroneous agent belief held by the system. Section 7 presents experimental results based on a real-world dataset. Section 8 finally concludes the paper.

2 SYSTEM MODEL & PROBLEM OVERVIEW

In this section, we introduce our system model in Section 2.1. In Section 2.2, we formulate our dynamic Bayesian game and give desired properties for mechanism design in Section 4.

2.1 System Model of Learning from Biased Agents

As shown in Figure 1, we consider a standard distributed learning process from $N \geq 2$ biased agents in two stages as follows:

- Stage I: Each selfish agent $i \in [N] := \{1, \cdots, N\}$ independently trains his local model parameter θ_i in bounded interval $[A_i, B_i]$ by using his local dataset $\mathcal{D}_i$.[1] Biased with his own loss function, he prefers the system to adopt his local model as the global one and may misreport another parameter $m_i(\theta_i)$ different from his actual θ_i to the system. The system and the other agents are uncertain of his θ_i realization. Instead, they can only infer from their historical training results that θ_i is roughly distributed according to a PDF $f_i(\cdot)$ with mean μ_i and variance σ_i^2. Later in Section 6, we extend our mechanism and design to address a more challenging case where the system holds erroneous belief of each agent i' s θ_i distribution.
- Stage II: After receiving N agents' messages $\{m_i(\theta_i)\}_{i=1}^N$, the system determines a global model $\hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)$ in the weighted-average $\sum_{i=1}^N w_i\theta_i$, where agent i's weight w_i is proportional to the size of his dataset $\mathcal{D}_i$ as $w_i = |\mathcal{D}_i|/(\sum_{j=1}^N |\mathcal{D}_j|)$ as in [5, 10], where $\sum_{i=1}^N w_i = 1$.

Each selfish agent i only cares about minimizing the square distance between his θ_i result and the system's global model decision $\hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)$, yet he is uncertain about the others' model parameters θ_{-i} and their messaging strategy $m_{-i}(\cdot)$ from their θ_{-i}. Similar to [10], we define his loss function $\ell_i(\theta_i, \hat{\theta}(\{m_j(\theta_j)\}_{j=1}^N))$ as the mean square error (MSE) as follows:

$$\begin{aligned}&\ell_i(\theta_i, \hat{\theta}(\{m_j(\theta_j)\}_{j=1}^N))\\=&\mathbb{E}_{\theta_{-i}, m_{-i}(\theta_{-i})}\big(\theta_i - \hat{\theta}(\{m_j(\theta_j)\}_{j=1}^N)\big)^2.\end{aligned} \tag{1}$$

Not biased on a particular agent's model, the system wants to fairly obtain the weighted model $\sum_{i=1}^N w_i\theta_i$ to represent all agents and regional data, and aims to minimize the MSE between the actual $\sum_{i=1}^N w_i\theta_i$ and its global inference decision $\hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)$ as follows:

$$\begin{aligned}&L\big(\hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)\big)\\=&\mathbb{E}_{\{\theta_i\}_{i=1}^N, \{m_i(\theta_i)\}_{i=1}^N}\left(\sum_{i=1}^N w_i\theta_i - \hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)\right)^2.\end{aligned} \tag{2}$$

2.2 Dynamic Bayesian Game Formulation

Based on our system model above, we formulate a dynamic Bayesian game to include each agent i's messaging $m_i(\theta_i)$ for minimizing (1) in Stage I and the system's inference for global model $\hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)$ to minimize (2) under incomplete information in Stage II. Note that the system has no access to any agent i's dataset and cannot verify if a biased agent's message is true ($m_i(\theta_i) = \theta_i$) or not. Further, biased agents are strategic in manipulating their messages against the system's inference for their own loss minimization, which makes it even harder for the system to learn their actual model parameters. For example, agent i expecting $\theta_i < \hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)$ will under-report his parameter $m_i(\theta_i) < \theta_i$ to mislead the global $\hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)$ to move closer to θ_i.

[1] θ_i can exist in either single or multiple dimensions. For the sake of clarity, we initially consider it to be single-dimensional. If it is multi-dimensional, we can still extend our method and analysis to each dimension independently.

To handle the above two challenges, we need to carefully design an information disclosure mechanism to incentivize each biased agent to message his model parameter. We should also ensure the final outcome of the global model to be as close to the actual $\sum_{i=1}^N w_i\theta_i$ as possible for efficiency of learning. We summarize these two desired properties for mechanism design as follows:

- *Truthfulness.* A mechanism is truthful if each agent i chooses the message $m_i(\theta_i)$ from his messaging space $\mathcal{M}_i$ that can reveal his model parameter θ_i as precise as possible.
- *Efficiency.* A mechanism is efficient if it reduces the system's MSE in (2) as much as possible.

Due to the page limit, we skip some detailed proofs but still introduce sketches in the maintext. Please refer to our online technical report [15] for details.

3 TWO BENCHMARK SCHEMES: BABBLING EQUILIBRIUM AND WEIGHTED MEDIAN

In this section, we analyze two common schemes used in the literature, which will serve as two benchmarks for our PAID mechanism to compare later.

3.1 Benchmark 1: Babbling Equilibrium

The current practice of the "babbling equilibrium" (e.g., [9, 11]) simply removes unverified agents' misleading updates as cheap-talks. At this equilibrium, the system blindly minimizes its MSE in (2) by determining the global model $\hat{\theta}$ according to probability distributions of $\{\theta_i\}_{i=1}^N$ as follows:

$$\min_{\hat{\theta}} \int_{A_N}^{B_N}\cdots\int_{A_1}^{B_1}\left(\sum_{i=1}^N w_i\theta_i - \hat{\theta}\right)^2 f_1(\theta_1)\cdots f_N(\theta_N)d\theta_1\cdots d\theta_N. \tag{3}$$

As (3) is convex in $\hat{\theta}$, we can solve it by the first-order condition to determine a global model $\hat{\theta}$. By substituting it into (2), we can obtain the MSE performance for the system.

Lemma 3.1. *At the babbling equilibrium of benchmark 1, the system determines the global model $\hat{\theta}$ as a weighted sum of the mean μ_i of each agent i's model parameter to obtain MSE:*

$$\hat{\theta} = \sum_{i=1}^N w_i\mu_i,\ L_1 = \sum_{i=1}^N w_i^2\sigma_i^2. \tag{4}$$

Besides, if each agent i has the same variance for local model parameter θ_i and the same weight $w_i = \frac{1}{N}$ for any $i \in [N]$ of equal-size local data, L_1 in (4) *decreases with the agent number N and tends to 0 only if $N \to \infty$.*

At benchmark 1, the system can never learn each agent i's model parameter θ_i. It is thus best to use the mean μ_i of each θ_i in the global model decision to minimize its MSE, which incurs a variance σ_i^2 from each $i \in [N]$. Besides, if each agent i has the same variance and weight, as N increases, more data inputs from other agents balance the difference between the weighted model $\sum_{i=1}^N w_i\theta_i$ and the weighted mean. The system thus incurs a smaller MSE from all the agents. Note that benchmark 1 needs a large number of agents for a small MSE, which is difficult to satisfy. We are thus motivated to find some other schemes to reduce the system's MSE.

3.2 Benchmark 2: Weighted Median Scheme

In the algorithmic game theory literature, the popular "weighted median" scheme is widely used to motivate strategic agents' truthful reporting, where the system commits with the weighted median of agents' messages (e.g., [7, 28]). We give the formal definition below.

Definition 3.2 (Weighted Median Scheme). The system first reorganizes agents' messages $\{m_i(\theta_i)\}_{i=1}^N$ in an increasing order as $m_{j_1} \leq \cdots \leq m_{j_N}$. It then chooses the weighted median $\hat{\theta} = m_{j_n}$ as its global decision, where

$$n = \min\left\{k \middle| \sum_{i<k} w_{j_i} \leq \frac{1}{2}, \sum_{i>k} w_{j_i} \leq \frac{1}{2}, 1 \leq k \leq N.\right\}.$$

Recall that each agent i's loss in (1) is singly-peaked at his model parameter θ_i. If his θ_i is less than weighted median m_{j_n}, his misreporting of $m_i < m_{j_n}$ does not alter m_{j_n}. If he misreports m_i to be larger than m_{j_n}, we have the new weighted median m'_{j_n} greater than m_{j_n}. His loss in (1) only increases due to $\theta_i < m_{j_n} \leq m'_{j_n}$. Similarly, he will not misreport if $\theta_i \geq m_{j_n}$. We have the following.

Lemma 3.3. *Each agent i will truthfully message his model parameter under the weighted median scheme, i.e.,* $m_i(\theta_i) = \theta_i, \forall i \in [N]$.

Now we are ready to analyze the system's MSE performance in benchmark 2. Substitute the weighted median parameter as the global model to (2), we derive the closed-form MSE under the case of uniformly distributed $\{\theta_i\}_{i=1}^N$ below. Though more involved, the analysis of non-uniform distributions is similar, and we examine the system's MSE for more general θ_i distributions later in Section 5.2.

Lemma 3.4. *Given each agent i's model parameter* $\theta_i \sim U[A, B]$ *for* $i \in [N]$*, the system's MSE in* (2) *under the weighted mean scheme is*

$$L_2 = \sum_{n=1}^{N}\sum_{m=1}^{N}\frac{1}{N}\sum_{m^+\subset\Phi_n}\frac{\mathbf{1}_{\sum_{i\in m^-} w_{j_i}\leq\frac{1}{2} \& \sum_{i\in m^+} w_{j_i}\leq\frac{1}{2}}}{C_{N-1}^{n-1}}\Bigg(\sum_{i\neq m}^{N} w_i^2\sigma^2$$
$$+\frac{3}{5}\sigma^2\Bigg(\sum_{i\neq m, i\in m^+} w_i \sum_{j\neq m,i; j\in m^+} w_j + \sum_{i\neq m, i\in m^-} w_i \sum_{j\neq m,i; j\in m^-} w_j\Bigg)$$
$$+\frac{1}{10}\sigma^2\Bigg(\sum_{i\neq m, i\in m^+} w_i \sum_{j\neq m,i; j\in m^-} w_j + \sum_{i\neq m, i\in m^-} w_i \sum_{j\neq m,i; j\in m^+} w_j\Bigg)\Bigg), \tag{5}$$

where $\mathbf{1}_{(\cdot)}$ *is the indicator function,* $m^+ = \{i|\theta_i > \theta_m\}$, $m^- = \{i|\theta_i \leq \theta_m, i \neq m\}$, $\Phi_n = \{m|\theta_m = \theta_{j_n}, m \in m^+\}$ *and* j_n *is the index of the* n_{th} *smallest parameter among* $\{\theta_i\}_{i=1}^N$. *If each agent i has model parameter* $\theta_i \sim U[A, B]$ *and identical weight* $w_i = \frac{1}{N}$ *for any* $i \in [N]$*, the system's MSE in* (5) *is simplified to*[2]

$$L_2 = \begin{cases} \frac{(N-1)(7N+1)}{20N^2}\sigma^2, & \text{if } N \text{ is odd,} \\ \frac{7N^2-6N+4}{20N^2}\sigma^2, & \text{if } N \text{ is even.} \end{cases}$$

L_2 *decreases as N increases from 2 to 3 and increases with* $N \geq 3$.

At benchmark 2, if each agent has the same weight and his θ_i has the same uniform distribution, the system incurs a smaller MSE as the agent number N increases from 2 to 3. The reason is that with $N = 2$, both agents are equally critical to the system's MSE due to having the same weight. The system's choice of either of them to be the global decision results in a non-small loss. However, as N keeps increasing from 3, there are more agents whose model parameters are different from the weighted median one. This causes the system to incur much loss with non-median agents according to (1). The system's MSE thus increases and cannot be zero even for $N \to \infty$, which is different from benchmark 1.

With Lemmas 3.1 and 3.4, we are now ready to compare the system's MSE under benchmarks 1 and 2 below for uniform distribution. We compare for more other θ_i distributions later in Section 5.2 to show similar results.

Corollary 3.5. *Given each agent i's model parameter* $\theta_i \sim U[A, B]$ *and the same weight* $w_i = \frac{1}{N}$ *for* $i \in [N]$*, the system in benchmark 2 incurs a smaller MSE than 1 if* $N \in \{2, 3\}$*, and incurs a larger MSE if* $N \geq 4$.

With a small agent number $N \in \{2, 3\}$, benchmark 2 incurs a smaller MSE for the system than benchmark 1 due to its agents' truthful reporting of model parameters. However, as N keeps increasing from 3 and there are more non-median agents, it incurs even a larger MSE than benchmark 1. We are well motivated to develop a brand new mechanism to substantially reduce the system's MSE for any agent number N and model distribution.

4 OUR NEW MECHANISM DESIGN: PAID

In this section, we design a PArtial Information Disclosure PAID mechanism to substantially decrease the system's MSE and will compare its performance against the two benchmarks later in Section 5.

4.1 Introduction of Our PAID Mechanism

According to our analysis of benchmark 2 in Section 3.2, it is too costly on the system side to ensure each agent i' truthful reporting of precise θ_i by committing the weighted median scheme (rather than the weighted mean). Instead, we no longer expect each agent to report precisely a point but a range, by relaxing the commitment for higher efficiency purpose. Note that each agent i's loss $\ell_i(\theta_i)$ in (1) is singly-peaked with θ_i, implying that he may truthfully message his θ_i realization belonging to a range. As a new messaging method, we propose to carefully partition θ_i's distribution range $[A_i, B_i]$ into K_i sub-intervals and ask agent i to tell which sub-interval his θ_i belongs to. We redefine such an agent i' messaging space $\mathcal{M}_i$ instead of $[A_i, B_i]$ in the following.

Definition 4.1 (Agents' Messaging Space M_i in PAID). The system determines Θ_i=$\{[\theta_{i,k-1}, \theta_{i,k}]\}_{k=1}^{K_i}$ as the set of K_i connected sub-intervals of agent i's θ_i distribution interval $[A_i, B_i]$, where $\theta_{i,0} = A_i$, $\theta_{i,K_i} = B_i$, and $\{\theta_{i,k}\}_{k=0}^{K_i}$ is a strictly increasing sequence in k. It then invites each agent i's messaging in space $\mathcal{M}_i = \{1, \cdots, K_i\}$, where agent i's message $m_i = k$ indicates $\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]$, $1 \leq k \leq K_i$.

As each agent i may have distinct $[A_i, B_i]$, we may partition differently. And the partitioned sub-intervals should be properly designed to ensure truthful messaging from each agent in M_i. Note that the babbling equilibrium of benchmark 1 is a special case of our PAID mechanism by setting $K_i = 1$ for any $i \in [N]$. Given Definition 4.1, we now introduce our PAID mechanism in the following.

[2] We choose the $\lfloor \frac{N}{2} \rfloor$th smallest message as the median of N messages.

Definition 4.2 (PArtial Information Disclosure Mechanism PAID). The system designs the new mechanism PAID as follows.

- Step 1: The system first determines the sub-interval number K_i and sub-interval set $\Theta_i = \{[\theta_{i,k-1},\theta_{i,k}]\}_{k=1}^{K_i}$ to each agent $i \in [N]$, where $\{\theta_{i,k}\}_{k=1}^{K_i-1}$ for any $K_i \geq 2$ are solutions to

$$\left(\theta_{i,k} - w_i \mathbb{E}[\theta_i | \theta_i \in [\theta_{i,k-1}, \theta_{i,k}]] - \sum_{j\neq i}^{N} w_j \mu_j\right)^2 \tag{6}$$
$$=\left(\theta_{i,k} - w_i \mathbb{E}[\theta_i | \theta_i \in [\theta_{i,k}, \theta_{i,k+1}]] - \sum_{j\neq i}^{N} w_j \mu_j\right)^2, 1 \leq k \leq K_i - 1,$$

where $\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]]$ denotes the conditional mean of θ_i given that it belongs to the sub-interval of $[\theta_{i,k-1}, \theta_{i,k}]$.
- Step 2: Each agent i chooses his message $m_i(\theta_i) = k, k \in \mathcal{M}_i$ to minimize his loss in (1), telling $\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]$.
- Step 3: After receiving agents' messages $\{m_i(\theta_i)\}_{i=1}^N$, the system determines the global model

$$\hat{\theta}^*(\{\hat{\theta}_i(m_i(\theta_i))\}_{i=1}^N) = \sum_{i=1}^N w_i \hat{\theta}_i(m_i(\theta_i)),$$
$$\hat{\theta}_i(m_i(\theta_i) = k) = \mathbb{E}[\theta_i | \theta_i \in [\theta_{i,k-1}, \theta_{i,k}]], 1 \leq k \leq K_i. \tag{7}$$

Note that (6) and K_i are properly designed. The next proposition shows that our PAID mechanism in Definition 4.2 ensures each agent i's truthful messaging of the sub-interval containing his θ_i.

Proposition 4.3. *Our PAID mechanism in Definition 4.2 is truthful such that each agent i truthfully messages:*

$$m_i^*(\theta_i) = \{k | \theta_i \in [\theta_{i,k-1}, \theta_{i,k}], 1 \leq k \leq K_i.\}.$$

4.2 Optimization of Our PAID Mechanism

To obtain clear engineering insights on our PAID mechanism, in the following, we first focus on the case of each agent i's uniformly distributed model parameter, i.e., $\theta_i \in U[A_i, B_i]$. We relax this assumption to analyze for any continuous θ_i distribution later in Section 5.2.

We use backward induction to solve our PAID mechanism. In Step 3, upon receiving agents' truthful messages $\{m_i(\theta_i)\}_{i=1}^N$ from Step 2, the system determines each $\hat{\theta}_i(m_i(\theta_i) = k) = (\theta_{i,k-1} + \theta_{i,k})/2$. Substituting $\hat{\theta}_i(m_i(\theta_i))$ into (6), we derive that agent i's sub-interval boundaries $\{\theta_{i,k}\}_{k=1}^{K_i-1}$ in Step 1 are the solutions to

$$\theta_{i,k+1} - \frac{2(2-w_i)}{w_i}\theta_{i,k} + \theta_{i,k-1} = -\frac{4\sum_{j\neq i} w_j \mu_j}{w_i}, k \leq K_i - 1. \tag{8}$$

As (8) is a linear, autonomous, second-order difference equation, we divide it into a steady-state part with $\theta_{i,k} = \bar{\theta}_i$ for all $k \geq 0$ and a homogeneous part:

$$\theta_{i,k+1} - \frac{2(2-w_i)}{w_i}\theta_{i,k} + \theta_{i,k-1} = 0, k \leq K_i - 1.$$

As the solution to (8) is the sum of the solutions to the steady-state and the homogeneous parts, we have the following.

Proposition 4.4. *Given each agent i's sub-interval number K_i fixed in Step 1 of our PAID mechanism, the system determines his sub-interval boundaries in closed-form $\{\theta_{i,k}\}_{k=0}^{K_i}$ as follows:*

$$\theta_{i,k} = \alpha_i(K_i)\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^k + \beta_i(K_i)\left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^k + \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i},\ 0 \leq k \leq K_i,\ i \in [N], \tag{9}$$

where

$$\alpha_i(K_i) = \frac{-\left(A_i - \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}\right)\left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i} + B_i - \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}}{\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}},$$

$$\beta_i(K_i) = \frac{\left(A_i - \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}\right)\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} + \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - B_i}{\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}}.$$

Note that we have not determined the sub-interval number K_i for each agent $i \in [N]$ yet. The system's objective is to choose the optimal $\{K_i^*\}_{i=1}^N$ to minimize its MSE in (2) subject to agents' truthful messaging. According to (2) and Proposition 4.4, we have the following closed-form result.

Corollary 4.5. *The system's MSE in* (2) *of our PAID is*

$$L^*(\{K_i\}_{i=1}^N) = \sum_{i=1}^N \Bigg\{ \frac{w_i(1-w_i)}{3(4-w_i)\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)^2}$$
$$\left(24\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - \mu_i\right)^2 + 6(A_i - B_i)^2 - 12\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - A_i\right)\right.$$
$$\left.\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - B_i\right)\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)\right)$$
$$+ \frac{w_i^2(1-w_i)}{3(4-w_i)}\left(3\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - \mu_i\right)^2 + \frac{1}{4}(A_i - B_i)^2\right)\Bigg\}, \tag{10}$$

which decreases with each sub-interval number K_i, $i \in [N]$.

With a larger sub-interval number K_i, the system divides more sub-intervals for agent i to report and can extract accurate information on his θ_i. For the special case of each agent i's sub-interval number $K_i = 1$ for $i \in [N]$, the system's MSE L^* in (10) degenerates to L_1 in (4) of benchmark 1.

According to Corollary 4.5, the system's MSE of our PAID decreases with each agent i's sub-interval number K_i. We thus design each K_i to be as large as possible to minimize the system's MSE. Note that the solutions in (9) cannot guarantee that $\{\theta_{i,k}\}_{k=0}^{K_i}$ is strictly increasing in k, which needs to be satisfied for ensuring each agent i's truthful messaging. We have the following.

Proposition 4.6. *In our PAID mechanism, the optimal sub-interval number K_i^* for each agent i to minimize its MSE in* (10) *is in closed-form, given by:*

$$K_i^* = \begin{cases} \left\lfloor \log_{\frac{(1+\sqrt{1-w_i})^2}{w_i}} \frac{C_s+\sqrt{C_s^2-4\frac{(1+\sqrt{1-w_i})^2}{w_i}}}{2} \right\rfloor, & \text{if } \bar{\mu}_{-i} \le A_i, \\ \infty, & \text{if } \bar{\mu}_{-i} \in (A_i, B_i), \\ \left\lfloor \log_{\frac{(1+\sqrt{1-w_i})^2}{w_i}} \frac{C_b+\sqrt{C_b^2-4\frac{(1+\sqrt{1-w_i})^2}{w_i}}}{2} \right\rfloor, & \text{if } \bar{\mu}_{-i} \ge B_i, \end{cases} \tag{11}$$

where we have

$$\bar{\mu}_{-i} = \frac{\sum_{j\neq i}^N w_j \mu_j}{1-w_i}, \; C_s = \left(\frac{(1+\sqrt{1-w_i})^2}{w_i}+1\right)\frac{B_i-\bar{\mu}_{-i}}{A_i-\bar{\mu}_{-i}},$$
$$C_b = \left(\frac{(1+\sqrt{1-w_i})^2}{w_i}+1\right)\frac{A_i-\bar{\mu}_{-i}}{B_i-\bar{\mu}_{-i}}.$$

For agent i to truthfully message, he considers the effect of the other agents' expected mean parameter μ_{-i} in the aggregation. If μ_{-i} is within his parameter region, they are similar and it is easier for the system to persuade the agent to follow large enough $K_i^* = \infty$ to reveal more local information. However, if μ_{-i} is outside his region, they are very different and agent i has the intention to correct his message to obviously affect the aggregation. As such, the system can only decide a finite K_i^* to ensure his truthfully message in the rough sub-intervals.

Corollary 4.7. *Each agent i's optimal sub-interval number K_i^* in* (11) *is non-increasing in his θ_i's lower bound A_i and is non-decreasing in the upper bound B_i, respectively. Besides, it increases with the others' expected mean $\bar{\mu}_{-i}$ in the range of $\bar{\mu}_{-i} < A_i$ and it decreases with $\bar{\mu}_{-i}$ in the range of $\bar{\mu}_{-i} > B_i$.*

With decreased A_i or increased B_i for a larger range $[A_i, B_i]$ for possible θ_i, the system tends to decide a larger sub-interval number K_i^* to narrow each sub-interval for better inference. As the others' expected mean $\bar{\mu}_{-i}$ increases to approach A_i, its disturbance on agent i's truthful messaging is mitigated. The system can divide more sub-intervals for agent i and K_i^* thus increases. However, as $\bar{\mu}_{-i}$ keeps increasing beyond B_i, its disturbance on agent i's truthful messaging becomes intensified. The system can only divide fewer sub-intervals for agent i and K_i^* thus decreases.

One may be curious about our PAID mechanism under $K_i^* \to \infty$ for $i \in [N]$ and wonders whether $K_i = \infty$ reveals the actual θ_i information. According to Propositions 4.4 and 4.6, we have the following to tell that it is not the case. This is necessary to ensure each agent's truthful messaging in the sub-intervals.

Proposition 4.8. *Given $\bar{\mu}_{-i} = \frac{\sum_{j\neq i}^N w_j\mu_j}{1-w_i}$ falls within (A_i, B_i) for any $i \in [N]$, the system optimally designs the sub-interval boundaries $\{\theta_{i,k}\}_{i=0}^{K_i^*=\infty}$ for PAID as $\theta_{i,0} = A_i$, $\theta_{i,\infty} = B_i$ and*

$$\theta_{i,k} = -(\bar{\mu}_{-i} - A_i)\left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^k + \bar{\mu}_{-i}, \; k < \infty, \tag{12}$$

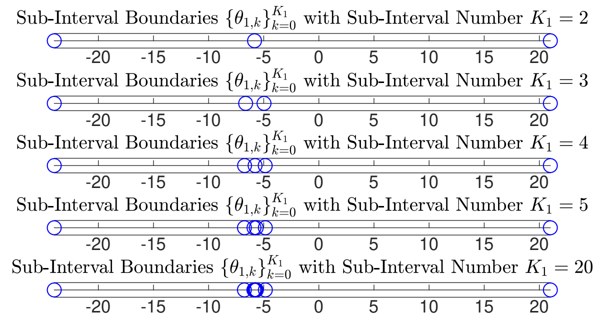

Figure 2: Agent 1's sub-interval boundaries $\{\theta_{1,k}\}_{k=0}^{K_1}$ in (9) **versus his sub-interval number K_1. Here we consider 4 agents in total and each agent i's θ_i follows a truncated normal distribution.**

which ensures the first and the last sub-intervals do not vanish with $K_i^* = \infty$:

$$\theta_{i,1} - \theta_{i,0} = (\bar{\mu}_{-i} - A_i)\left(1 - \frac{(1-\sqrt{1-w_i})^2}{w_i}\right) > 0,$$
$$\lim_{K_i^*\to\infty} \theta_{i,K_i^*} - \theta_{i,K_i^*-1} = B_i - \bar{\mu}_{-i} > 0.$$

Figure 2 considers an example of truthful learning from 4 biased agents, with each θ_i following a truncated normal distribution. Consider a particular agent 1, as his K_1 increases, the system divides more sub-intervals around the mean -6 in the range of [-24, 21]. Yet, the first and the last sub-intervals $[\theta_{1,0}, \theta_{1,1}]$ and $[\theta_{1,K_1-1}, \theta_{1,K_1}]$ do not vanish as K_1 increases and attains the value 20, which is consistent with Proposition 4.8. This ambiguity helps hide agent i's actual θ_i and ensures his truthful messaging of θ_i's sub-interval.

5 PAID PERFORMANCE ANALYSIS AND COMPARISON WITH TWO BENCHMARKS

Besides truthful property, in this section, we continue to show our PAID's efficiency advantages over the benchmark schemes in Section 3 to reduce the system's MSE. We first make analytical comparisons in Section 5.1. In Section 5.2, we propose an efficient algorithm for our PAID design to fit more general θ_i distributions and provide numerical comparison with the two benchmarks.

5.1 Analytical Comparisons with Benchmarks

Combining Corollary 4.5 and Proposition 4.6, we can obtain the system's MSE in (2) of our PAID mechanism in the following.

Lemma 5.1. *Given each agent i's model parameter $\theta_i \in U[A_i, B_i]$ with the same mean $\mu_i = \mu$ for any $i \in [N]$, the system's MSE in* (10) *of our PAID mechanism is*

$$L^* = \sum_{i=1}^N \frac{(1-w_i)w_i^2}{4-w_i}\sigma_i^2. \tag{13}$$

Specifically, if each agent i has the parameter $\theta_i \sim U[A, B]$ in the same interval and has the same weight $w_i = \frac{1}{N}$ for any $i \in [N]$, L^* in (13) *decreases with $N \ge 2$ and tends to 0 as $N \to \infty$.*

Algorithm 1 Solving sub-interval boundaries $\{\{\theta_{i,k}\}_{k=0}^{K_i}\}_{i=1}^N$ in (6) for general $\{\theta_i\}_{i=1}^N$ distributions.

Input: Agent number N, agents' parameter realization ranges $\{[A_i, B_i]\}_{i=1}^N$, PDFs $\{f_i(\cdot)\}_{i=1}^N$, agents' maximum searching rounds $\{T_i\}_{i=1}^N$, agents' parameter means $\{\mu_i\}_{i=1}^N$, objective functions $\{\mathbf{h}_i\}_{i=1}^N$, and digit precisions $\{\epsilon_i\}_{i=1}^N$.

Output: Sub-interval boundaries $\{\{\theta_{i,k}\}_{k=0}^{K_i^{max}}\}_{i=1}^N$.

1: **Initialization**: $K_i \leftarrow 2, \mathbf{s}_i^0 \leftarrow (s_{i,1}^0 \cdots, s_{i,K_i-1}^0), t \leftarrow 0, \mathbf{\Gamma}_1 \leftarrow \emptyset$, $\theta_{i,0} \leftarrow A_i$, $\theta_{i,K_i} \leftarrow B_i$.
2: Update $\mathbf{s}_i^{t+1} \leftarrow \mathbf{s}_i^t - \mathbf{J}_i(\mathbf{s}_i^t)^{-1} * \mathbf{h}_i(\mathbf{s}_i^t)$ until $t \geq T_i$ or $||\mathbf{s}_i^t - \mathbf{s}_i^{t-1}||_2 \leq 10^{-\epsilon_i}$. Use the pseudo-inverse of $\mathbf{J}_i(\mathbf{s}_i^t)$ if $\mathbf{J}_i(\mathbf{s}_i^t)^{-1}$ does not exist.
3: **if** $\exists t < T_i$ such that $||\mathbf{s}_i^{t+1} - \mathbf{s}_i^t||_2 \leq 10^{-\epsilon_i}$ **then**
4: Record $\mathbf{\Gamma}_{K_i} \leftarrow \mathbf{s}_i^{t+1}$ and repeat Steps 1-3 with $K_i \leftarrow K_i + 1$.
5: **else**
6: Determine $K_i^{max} \leftarrow K_i - 1$, $\{\theta_{i,k}\}_{k=0}^{K_i^{max}} \leftarrow (A_i, \mathbf{\Gamma}_{K_i^{max}}, B_i)$.
7: Repeat Steps 1-6 for all the agent $i \in [N]$.

In the following, we first compare the system's MSE of our PAID mechanism with benchmark 1.

Proposition 5.2. *Given each agent i has the model parameter $\theta_i \in U[A_i, B_i]$ with the same mean $\mu_i = \mu$ for any $i \in [N]$, the ratio r_1 between the system's MSE L^* in* (13) *of our PAID mechanism and L_1 in* (4) *of benchmark 1 is*

$$r_1 = \frac{L^*}{L_1} = \frac{\sum_{i=1}^N \frac{(1-w_i)w_i^2}{4-w_i}\sigma_i^2}{\sum_{j=1}^N w_j^2 \sigma_j^2}. \tag{14}$$

r_1 is always less than $\frac{1}{4}$. Besides, if each agent i has the model parameter $\theta_i \sim U[A, B]$ and has the same weight $w_i = \frac{1}{N}$ for any $i \in [N]$, we have r_1 in (14) *increases with N and tends to $\frac{1}{4}$ as $N \to \infty$.*

Our PAID mechanism substantially reduces the system MSE as compared to the benchmark 1 for any N. As benchmark 1 performs its best as $N \to \infty$, the ratio r_1 reaches 1/4 then.

Now we compare the system's MSE of our PAID with benchmark 2. We have the following.

Proposition 5.3. *If each agent i has the model parameter $\theta_i \sim U[A, B]$ and the same weight $w_i = \frac{1}{N}$ for any $i \in [N]$, the ratio r_2 between the system's MSE L^* in* (13) *of our PAID and L_2 in* (5) *of benchmark 2 is*

$$r_2 = \frac{L^*}{L_2} = \begin{cases} \frac{20N}{(4N-1)(7N+1)}, & \text{if } N \text{ is odd,} \\ \frac{20(N-1)N}{(4N-1)(7N^2-6N+4)}, & \text{if } N \text{ is even.} \end{cases} \tag{15}$$

r_2 is no more than $\frac{2}{7}$, decreases with N and tends to 0 as $N \to \infty$.

As the agent number N increases, there are more agents whose parameter realizations are different from the weighted median and benchmark 2 worsens as shown in Lemma 3.4. This causes the ratio r_2 to decrease with N and our PAID mechanism's advantage over benchmark 2 becomes more obvious.

5.2 Performance Comparison for Other θ_i Distributions

Recall that in Sections 4.2 and 5.1, we assume that each θ_i follows a uniform distribution. In this subsection, we relax this assumption to fit our PAID for each generally distributed θ_i. We first give a sufficient condition for agent i's sub-interval number K_i design below.

Lemma 5.4. *Given any continuous PDF $f_i(\cdot)$ of agent i's model parameter $\theta_i \in [A_i, B_i]$, his sub-interval number K_i in our PAID is no smaller than 2 as long as the following condition is satisfied:*

$$\frac{(2-w_i)A_i - w_i\mu_i}{2} < \sum_{j=1, j\neq i}^N w_j\mu_j < \frac{(2-w_i)B_i - w_i\mu_i}{2}.$$

Lemma 5.4 is similar to Proposition 4.6 for the case of uniformly distributed θ_i. Recall that $K_i = 1$ reveals no information and degenerates to benchmark 1. If the others' weighted mean $\sum_{j=1, j\neq i}^N w_j\mu_j$ lies in a range close to agent i's θ_i realization range of $[A_i, B_i]$, each agent i worries less about the disturbance from the other agents, and is easier to be persuaded by the system to message truthfully in more partitioned sub-intervals to reveal local information.

In the following, we introduce an efficient algorithm in Algorithm 1 to solve each agent i's truthful sub-interval boundaries $\{\theta_{i,k}\}_{k=0}^{K_i}$ for any θ_i distribution by solving a set of equations in (6) and determining the maximum possible K_i called K_i^{max}. Let us first introduce some essential notations. Define vectors $\mathbf{s}_i^j = (s_{i,1}^j, \cdots, s_{i,K_i-1}^j)$ and $\mathbf{x}_i = (x_{i,1}, \cdots, x_{i,K_i-1})$. Define the objective functions $\mathbf{h}_i := (h_{i,1}, \cdots, h_{i,K_i-1})$ such that (6) can be rewritten as $h_{i,k}(\theta_{i,k}) = 0$. Denote $\mathbf{J}_i(\mathbf{s})$ as the Jacobian matrix of $\mathbf{h}_i(\mathbf{s})$. Our objective is to numerically find a solution $\mathbf{s}$ of $\mathbf{h}_i$ such that $\mathbf{h}_i(\mathbf{s}) = 0$. Inspired by Corollary 4.5, for better system performance, we design Algorithm 1 to initiate with a sub-interval number of $K_i = 2$ and stop when monotonicity of $\theta_{i,k}$ sequence almost fails to obtain the largest feasible K_i^{max} that supports truthful messaging for agent i. The computational complexity of Algorithm 1 is

$$O\left(\sum_{i=1}^N \sum_{K_i=2}^{K_i^{max}} K_i^3 \cdot \log_2 \epsilon_i\right),$$

where N denote the agent number, K_i^{max} denotes agent i's maximum possible sub-interval number, and ϵ_i is the digit precision of agent i. Thus, Algorithm 1 incurs a polynomial complexity order in each agent i's maximum possible sub-interval number K_i^{max}. It is also scalable with large agent number N. Additionally, Algorithm 1 exhibits good convergence with a quadratic rate.

Lemma 5.5. *Suppose that agent i's model parameter θ_i follows a truncated normal distribution. Denote $\boldsymbol{\theta}_i^*$ is a root of $\mathbf{h}_i$ in Algorithm 1, such that $\mathbf{h}_i(\boldsymbol{\theta}_i^*) = 0$. There exists a positive $\delta > 0$ such that if $||\mathbf{s}_i^0 - \boldsymbol{\theta}_i^*|| < \delta$ holds, we have*

$$\lim_{t\to\infty} \frac{||\mathbf{s}_i^{t+1} - \boldsymbol{\theta}_i^*||}{||\mathbf{s}_i^t - \boldsymbol{\theta}_i^*||} = 0.$$

Besides, there exists a positive $M > 0$ such that for all $t \geq 0$,

$$||\mathbf{s}_i^{t+1} - \boldsymbol{\theta}_i^*|| \leq M||\mathbf{s}_i^t - \boldsymbol{\theta}_i^*||^2.$$

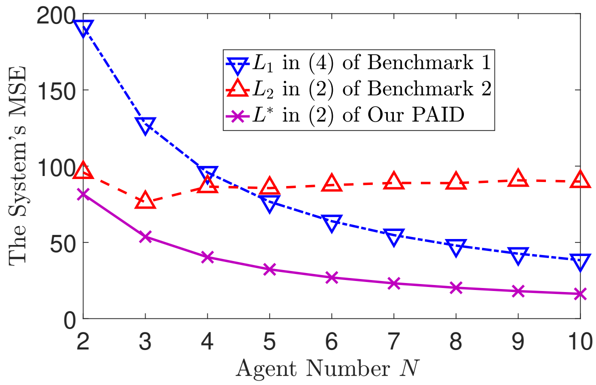

Figure 3: The system's MSE L_1 in (4) of benchmark 1, L_2 in (2) of benchmark 2 and L^* in (2) of our PAID versus the agent number N. Here we consider that each agent i's parameter θ_i follows a truncated normal distribution and has the same weight of $w_i = \frac{1}{N}$.

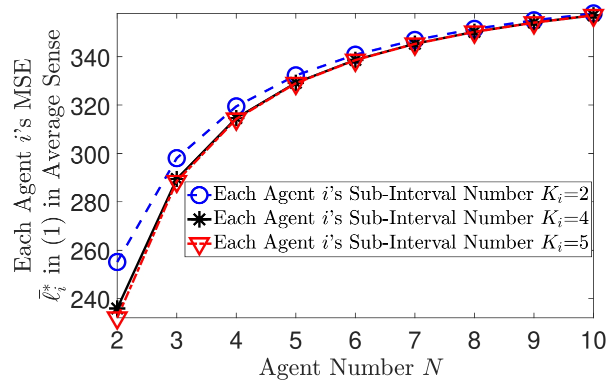

Figure 4: Each agent i's MSE $\bar{\ell}_i$ in (1) in average sense of our PAID versus the agent number N and each agent i's sub-interval number K_i, respectively. Here we consider that each agent i's parameter θ_i follows a truncated normal distribution and has the same weight of $w_i = \frac{1}{N}$.

In the following, we numerically analyze and compare the performance of our PAID and the two benchmarks for truncated normal distributions of agents' model parameters.

Figure 3 shows the system's MSE L_1 in (4) of benchmark 1, L_2 in (2) of benchmark 2 and L^* in (2) of our PAID versus the agent number N, respectively. Here we consider each agent has the same weight $w_i = \frac{1}{N}$. Figure 3 shows that the system always incurs a much smaller MSE under our PAID than the two benchmarks, which is consistent with Propositions 5.2 and 5.3. Besides, the system first incurs a larger MSE under benchmark 1 than 2 if the agent number is small as $N \le 4$, and then incurs a smaller MSE if $N \ge 5$, which is consistent with Corollary 3.5.

Figure 4 shows each agent i's MSE in (1) in average sense of our PAID versus the agent number N and agents' sub-interval number K, respectively. As N increases, each agent expects more uncertainty of the others' model parameter realizations and his MSE then increases. Perhaps surprisingly, like the system error in Corollary 4.5, each agent i's MSE also decreases with his sub-interval number K_i, the system can learn each agent i's θ_i more precisely, and in return the local model of the agent is better reflected in the global model. Figure 4 with minor difference for larger K_i also implies that the system does not need to implement too many sub-intervals due to the diminishing return over K_i.

6 EXTENSION TO INCLUDE ERRONEOUS AGENT BELIEF AT THE SYSTEM

Recall that in Section 2, we assume that the system perfectly knows the probability distribution $f_i(\cdot)$ of each agent i's model parameter $\theta_i \in [A_i, B_i]$, $i \in [N]$. In this section, we extend our PAID to address the more challenging scenario where the system holds erroneous belief of each agent i's model parameter distribution. After imposing an error belief to the system, now θ_i's distribution changes from a deterministic range $[A_i, B_i]$ to $[A_i - n_i, B_i + n_i]$, where $n_i \in [-D, D]$ is the random noise and is unknown to the system. We still suppose $A_i + B_i = 2\mu_i$ and bound the maximum deviation $D \in (\frac{A_i - B_i}{2}, \frac{B_i - A_i}{2})$. We focus on the case that PDF $f_i(\cdot)$ of each θ_i is symmetric, defined below.

Definition 6.1. A PDF $f_i(\cdot)$ is defined as symmetric if it satisfies $f(\mu_i - x) = f(\mu_i + x)$ for any $x \in \mathcal{R}$, where μ_i is the mean.

Such erroneous belief n_i on $[A_i - n_i, B_i + n_i]$ introduces difficulty to our prior PAID's partition into sub-intervals. Our PAID design is still to incentivize agents' truthful messaging of the sub-intervals containing their model parameters. Our basic idea is to determine each agent i's sub-interval boundaries $\{\theta_{i,k}\}_{k=0}^{K_i}$ of our PAID according to his θ_i's largest possible range of $[A_i - D, B_i + D]$. By refining our analysis in Section 4, we have the following.

Proposition 6.2. *Suppose that each agent i's model parameter θ_i follows a symmetric PDF $f_i(\cdot)$ in a range of $[A_i - n_i, B_i + n_i]$ with mean μ_i and unknown $n_i \in [-D, D]$ for $i \in [N]$. The system optimally determines each agent i's sub-interval boundaries to be $\theta_{i,0} = A_i - D$, $\theta_{i,K_i} = B_i + D$ and $\{\theta_{i,k}\}_{k=1}^{K_i^*-1}$ as the solutions to*

$$(\theta_{i,k} - w_i E[\theta_i | \theta_i \in [\theta_{i,k-1}, \theta_{i,k}]] - \sum_{j \neq i} w_j \mu_j)^2$$
$$= (\theta_{i,k} - w_i E[\theta_i | \theta_i \in [\theta_{i,k}, \theta_{i,k+1}]] - \sum_{j \neq i} w_j \mu_j)^2, \; k \le K_i^* - 1,$$

where K_i^* is optimally decided as agent i's largest sub-interval number such that $\{\theta_{i,k}\}_{i=0}^{K_i^*}$ is a strictly increasing sequence. Each agent i will truthfully message the index of the sub-interval containing his θ_i:

$$m_i^*(\theta_i) = \{k | \theta_i \in [\theta_{i,k-1}, \theta_{i,k}], k \in \{1, \cdots, K_i^*\}.\}.$$

To tell the effect of the system's erroneous belief, we define the system's MSE increment ΔL^* caused by uncertain bounds of agent i's θ_i as follow:

$$\Delta L^* = \mathbb{E}_{\{\theta_i\}_{i=1}^N, \{n_i\}_{i=1}^N}[L^*] - \mathbb{E}_{\{\theta_i \in [A_i, B_i]\}_{i=1}^N}[L^* | n_i = 0]. \quad (16)$$

The following proposition indicates that ΔL^* is limited in a quadratic term of D.

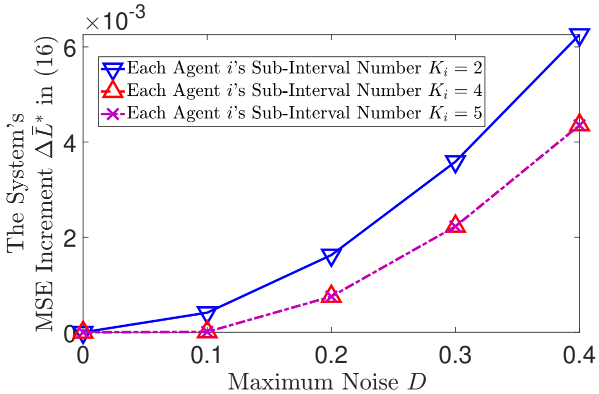

Figure 5: The system's MSE increment ΔL^* in (16) versus the maximum noise D and agent i's identical sub-interval number K_i, respectively. Here we consider $N = 6$ agents in total and each agent i's θ_i follows a truncated normal distribution.

Proposition 6.3. *Given each agent i's model parameter $\theta_i \sim U[A_i - n_i, B_i + n_i]$ and the noise $n_i \sim U[-D, D]$ for any $i \in [N]$, the system's MSE increment ΔL^* in (16) has the limited loss of $O(D^2)$.*

Figure 5 shows the system's MSE increment ΔL^* in (16) versus the maximum noise D and agent i's sub-interval number K_i, respectively. It indicates that the system's MSE increment grows as the maximum noise D increases from 0 to 0.4 but the maximum loss is slightly greater than 6×10^{-3}, which is consistent with Proposition 6.3. Further, the system incurs a smaller MSE increment as the sub-interval number increases from $K_i = 2$ to 5, which is consistent with Corollary 4.5 and our K_i^* design in Proposition 6.2.

7 CASE STUDY: SURVIVAL PREDICTION FROM REGIONAL HOSPITALS

In this section, we evaluate our PAID's performance using a real-world dataset. The system (National Cancer Institute) seeks to develop and deploy a learning model to predict BReast CAncer (BRCA) patients' overall survival days from their ages at diagnosis (e.g., [26]), which will used by all the hospitals (i.e., agents). The dataset on BRCA study is obtained from the cancer genome Atlas's genomics data commons portal [21], which contains patients' ages at diagnosis and overall survival days of 875 patients in the USA.

The system asks hospitals of different geographical regions (northeast, south, west, middlewest) to train a linear regression model as follows:

$$\mathbf{y}_i = \theta_i \mathbf{x}_i + y_0 \cdot \mathbf{1}, i = 1, \cdots, 4, \tag{17}$$

where $\mathbf{y}_i$ as a column vector denotes overall survival days of hospital i's patients, θ_i is hospital i's biological significance parameter, $\mathbf{x}_i$ is the column vector of ages at diagnosis of hospital i's patients, and y_0 is a constant for 5-year survival baseline (e.g., [12, 13]). Each hospital i obtains its parameter θ_i using ordinary least squares method. The system obtains the global model $\hat{\theta}$ of our PAID according to (7). More details for our experiment settings and processing are given in our online technical report [15].

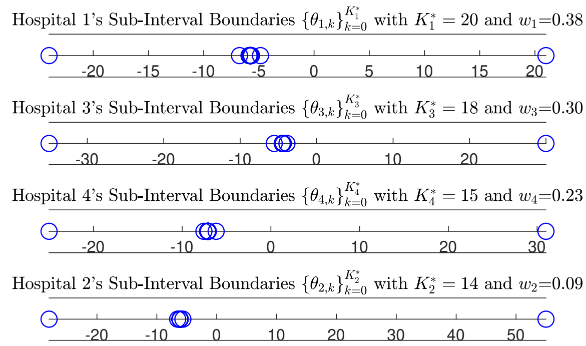

Figure 6: Each hospital i's sub-interval boundaries $\{\theta_{i,k}\}_{k=0}^{K_i^*}$ in (6) of our PAID mechanism, according to its maximum sub-interval number K_i^* and weight w_i, $i = 1, \cdots, 4$.

Figure 6 shows that each hospital i's sub-interval boundaries $\{\theta_{i,k}\}_{k=0}^{K_i^*}$ in (6) with sub-interval number K_i^* of our PAID returned by Algorithm 1. Note that each hospital i has $K_i^* - 2$ sub-intervals around its distinct θ_i mean, and its first and last sub-intervals are the longest, which is consistent with Proposition 4.8. We also note that hospital 1 with the largest weight $w_1 = 0.38$ has the largest sub-interval number $K_1 = 20$ since the system views it most importantly and is eager to extract more precise information from more partitions of the interval [-24, 21].

Similar to (2), the system's actual MSE for survival prediction is empirically calculated over all hospital datasets $\{\mathcal{D}_i\}_{i=1}^4$ using the global model decision $\hat{\theta}$ as follows:

$$MSE(\hat{\theta}) = \frac{1}{\sum_{k=1}^4 |\mathcal{D}_k|} \sum_{i=1}^4 \sum_{(x_j, y_j) \in \mathcal{D}_i} (y_j - \hat{\theta} x_j - y_0)^2. \tag{18}$$

Similar to (1), each hospital i incurs a different MSE from the global model as follows:

$$MSE_i(\hat{\theta}) = \frac{1}{|\mathcal{D}_i|} \sum_{(x_j, y_j) \in \mathcal{D}_i} (y_j - \hat{\theta} x_j - y_0)^2, \; i = 1, \cdots, 4. \tag{19}$$

We define the system's MSE increment as the difference between its MSE in (18) of global decision $\hat{\theta}$ and that of the actual $\hat{\theta}^{**} = \sum_{i=1}^N w_i \theta_i$ as follows:

$$\Delta MSE(\hat{\theta}) = MSE(\hat{\theta}) - MSE(\hat{\theta}^{**}), \tag{20}$$

and define a hospital i's MSE increment as the difference between its MSE in (19) of global decision $\hat{\theta}$ and that of the actual $\hat{\theta}^{**} = \sum_{i=1}^N w_i \theta_i$ as follows:

$$\Delta MSE_i(\hat{\theta}) = MSE_i(\hat{\theta}) - MSE_i(\hat{\theta}^{**}), i = 1, \cdots, 4. \tag{21}$$

Figure 7 compares the system's MSE increment in (20) under our PAID against two benchmarks. Figure 7 also shows each hospital's MSE increment in (21). It indicates that our PAID always incurs a smaller MSE for the system than two benchmarks. We also find that our PAID incurs a smaller MSE for most hospitals (i.e., 1, 2, 4) than the two benchmarks. Recall in benchmark 1, the system

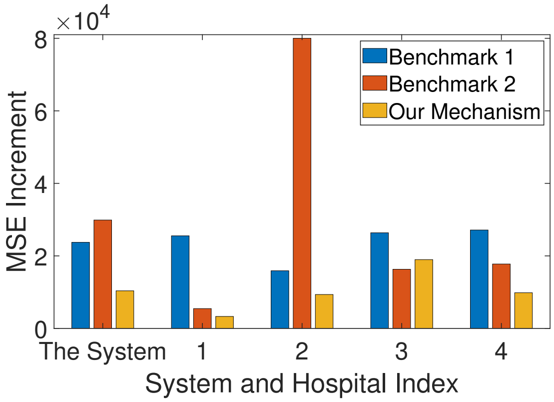

Figure 7: The system's MSE increment in (20) **and each hospital's MSE increment in** (21) **under benchmarks 1, 2, and our PAID, respectively.**

blindly averages similar means of four hospitals' parameters as the global model. Therefore, each hospital's MSE increment does not vary much from the others'. At benchmark 2, the system here chooses hospital 1's local model parameter as the global model. This leads to its small MSE by sacrificing all the others', especially for hospital 2. Differently, our PAID learns each hospital's parameter realization as precisely as possible and fairly averages them in the global decision. Each hospital thus obtains a smaller MSE increment than benchmark 1 and most hospitals also obtain smaller MSE increment than benchmark 2.

8 CONCLUSION

In this paper, we study how to truthfully learn from biased agents for training a fair global model for all agents. We find the current practice of the "babbling equilibrium" performs poorly. Though the popular "weighted median" scheme in the algorithmic game theory literature ensures agents' truthful reporting, we prove that it may perform even worse than the babbling equilibrium. Alternatively, we design a PAID mechanism to truthfully learn from biased agents while minimizing the global learning error. We prove that PAID's learning error is at most 1/4 of the babbling equilibrium and 2/7 of the weighted median scheme, respectively. We also show that our PAID is robust to erroneous agent belief held by the system to yield bounded loss. Finally, we run experiments using a real-world dataset to demonstrate our PAID's great advantages over the other approaches and show its positive "side-effect" to even benefit most biased agents.

REFERENCES

[1] Annie Abay, Yi Zhou, Nathalie Baracaldo, Shashank Rajamoni, Ebube Chuba, and Heiko Ludwig. 2020. Mitigating bias in federated learning. *arXiv preprint arXiv:2012.02447* (2020).

[2] Emanuel Scoullos Annika Brundyn and David Williams. 2022. Using Federated Learning to Bridge Data Silos in Financial Services. Nvidia Technical Blog. https://developer.nvidia.com/blog/using-federated-learning-to-bridge-data-silos-in-financial-services/#

[3] Aporia. Accessed in Mar. 2024. Dynamic Pricing models: Types, algorithms, and best practices. https://www.aporia.com/blog/learn/machine-learning-for-business/dynamic-pricing-models-types-algorithms-and-best-practices/

[4] Raphael Boleslavsky and Tracy R Lewis. 2016. Evolving influence: Mitigating extreme conflicts of interest in advisory relationships. *Games and economic behavior* 98 (2016), 110–134.

[5] Michael Borenstein, Larry V Hedges, Julian PT Higgins, and Hannah R Rothstein. 2021. *Introduction to meta-analysis.* John Wiley & Sons.

[6] Djamila Bouhata, Hamouma Moumen, Jocelyn Ahmed Mazari, and Ahcène Bounceur. 2022. Byzantine fault tolerance in distributed machine learning: a survey. *arXiv preprint arXiv:2205.02572* (2022).

[7] Hau Chan, Aris Filos-Ratsikas, Bo Li, Minming Li, and Chenhao Wang. 2021. Mechanism design for facility location problems: a survey. *arXiv preprint arXiv:2106.03457* (2021).

[8] Alexandre Chausson, EA Welden, Marina S Melanidis, Erin Gray, Mark Hirons, and Nathalie Seddon. 2023. Going beyond market-based mechanisms to finance nature-based solutions and foster sustainable futures. *PLOS climate* 2, 4 (2023), e0000169.

[9] Ying Chen, Navin Kartik, and Joel Sobel. 2008. Selecting cheap-talk equilibria. *Econometrica* 76, 1 (2008), 117–136.

[10] Kate Donahue and Jon Kleinberg. 2021. Model-sharing games: Analyzing federated learning under voluntary participation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 5303–5311.

[11] Joseph Farrell and Matthew Rabin. 1996. Cheap talk. *Journal of Economic perspectives* 10, 3 (1996), 103–118.

[12] Juliana Oliveira Fernandes, Cassio Cardoso-Filho, Maria Beatriz Kraft, Amanda Sacilotto Detoni, Barbara Narciso Duarte, Julia Yoriko Shinzato, and Diama Bhadra Vale. 2023. Differences in breast cancer survival and stage by age in off-target screening groups: a population-based retrospective study. *AJOG Global Reports* 3, 2 (2023), 100208.

[13] Sophie Frank, Matthieu Carton, Coraline Dubot, Mario Campone, Barbara Pistilli, Florence Dalenc, Audrey Mailliez, Christelle Lévy, Veronique d'Hondt, Marc Debled, et al. 2020. Impact of age at diagnosis of metastatic breast cancer on overall survival in the real-life ESME metastatic breast cancer cohort. *The Breast* 52 (2020), 50–57.

[14] Shugang Hao and Lingjie Duan. 2024. To Save Mobile Crowdsourcing from Cheap-talk: A Game Theoretic Learning Approach. *IEEE Transactions on Mobile Computing* (2024).

[15] S. Hao and L. Duan. 2024. Truthful Learning from Biased Agents. Techincal Report. https://shuganghao.github.io/files/hao2024.pdf

[16] Michael Hoy, John Livernois, Chris McKenna, Ray Rees, and Thanasis Stengos. 2022. *Mathematics for economics.* MIT press.

[17] Lutz Kaufmann and Soenke Roessing. 2005. Managing conflict of interests between headquarters and their subsidiaries regarding technology transfer to emerging markets—a framework. *Journal of World Business* 40, 3 (2005), 235–253.

[18] Bo Liu, Ming Ding, Sina Shaham, Wenny Rahayu, Farhad Farokhi, and Zihuai Lin. 2021. When machine learning meets privacy: A survey and outlook. *ACM Computing Surveys (CSUR)* 54, 2 (2021), 1–36.

[19] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. 2021. *Introduction to linear regression analysis.* John Wiley & Sons.

[20] Kristina Müller and Freimut Bodendorf. 2023. Cross-Silo Federated Learning in Enterprise Networks With Cooperative and Competing Actors. *The Human Side of Service Engineering* 108, 108 (2023).

[21] TCGA Research Network. [n. d.]. Tensorflow federated stack overflow dataset. https://www.cancer.gov/ccg/research/genome-sequencing/tcga

[22] Jorge Nocedal and Stephen J Wright. 1999. *Numerical optimization.* Springer.

[23] Hiromasa Ogawa. 2021. Receiver's sensitivity and strategic information transmission in multi-sender cheap talk. *International Journal of Game Theory* 50, 1 (2021), 215–239.

[24] Saurav Prakash and Amir Salman Avestimehr. 2020. Mitigating byzantine attacks in federated learning. *arXiv preprint arXiv:2010.07541* (2020).

[25] Ming Tang and Vincent WS Wong. 2023. Tackling system induced bias in federated learning: Stratification and convergence analysis. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications.* IEEE, 1–10.

[26] Jean Ogier du Terrail, Samy-Safwan Ayed, Edwige Cyffers, Felix Grimberg, Chaoyang He, Regis Loeb, Paul Mangold, Tanguy Marchand, Othmane Marfoq, Erum Mushtaq, et al. 2022. Flamby: Datasets and benchmarks for cross-silo federated learning in realistic healthcare settings. *arXiv preprint arXiv:2210.04620* (2022).

[27] Bassey Ude, Olusola T Odeyomi, Kaushik Roy, and Xiaohong Yuan. 2023. A Survey on Bias Mitigation in Federated Learning. In *2023 IEEE Symposium Series on Computational Intelligence (SSCI).* IEEE, 1170–1175.

[28] Xinping Xu, Lingjie Duan, and Minming Li. 2019. Strategic learning approach for deploying UAV-provided wireless services. *IEEE Transactions on Mobile Computing* 20, 3 (2019), 1230–1241.

[29] Yuan-Yi Xu, Ci-Siang Lin, and Yu-Chiang Frank Wang. 2023. Bias-Eliminating Augmentation Learning for Debiased Federated Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 20442–20452.

[30] Caiyi Yang and Javad Ghaderi. 2024. Byzantine-Robust Decentralized Learning via Remove-then-Clip Aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence.*

[31] Dong Yin, Yudong Chen, Ramchandran Kannan, and Peter Bartlett. 2018. Byzantine-robust distributed learning: Towards optimal statistical rates. In *International Conference on Machine Learning*. PMLR, 5650–5659.

Table 1: Information of hospitals' parameters and weights

No.	Mean	Variance	Range	Weight
1	-6.1	4.8^2	[-24 21]	0.38
2	-4.3	8.6^2	[-28 55]	0.09
3	-9.7	5.6^2	[-35 30]	0.30
4	-1.5	6.3^2	[-25 31]	0.23

A SET UP OF CASE STUDY IN SECTION 7

We follow [26] to group patients according to geographic regions to obtain sufficient data for each agent, conceptualized as a regional hospital. This grouping results in $N = 4$ regional hospitals across the USA: USA Northeast, USA South, USA West, and USA Middlewest, holding 311, 196, 206, and 162 patients.

In large samples, the distribution of a linear regression model parameter's estimates tends to approximate a normal distribution according to the central limit theorem (e.g., [19]). We then model that each θ_i follows a truncated normal distribution $\theta_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ in the range of $[A_i, B_i]$. To obtain these parameters, we randomly divide each hospital i's dataset into many batches $\mathcal{B}_j$ of size 10 and calculate each batch j's linear regression parameter according to (17), using ordinary least squares (OLS) method as follows:

$$\theta_j(\mathcal{B}_j) = \frac{\sum_{(x_k, y_k)\in\mathcal{B}_j} x_k(y_k - y_0)}{\sum_{(x_k, y_k)\in\mathcal{B}_j} x_k^2}. \tag{22}$$

We repeat the dividing and calculating process on each hospital's dataset for 10^4 rounds and take the average on the obtained parameters to approximate the mean, variance, and boundaries. Each hospital i's weight $w_i = (1/\sigma_i^2)/(\sum_{j=1}^N 1/\sigma_j^2)$ is proportional to his inverse variance $1/\sigma_i^2$ (e.g., [5]). We summarize in Table 1.

To obtain each hospital i's θ_i realization, we randomly select 10 patients from its dataset to formulate a testing dataset $\mathcal{T}_i$ to calculate the linear regression model parameter $\theta_i(\mathcal{T}_i)$ according to (22). We are then able to obtain the system's global model decision $\hat{\theta}$ under benchmarks 1, 2 and our mechanism according to $\{\theta_i(\mathcal{T}_i)\}_{i=1}^4$.

We obtain the system's MSE in Figure 7 as the mean of $MSE(\hat{\theta})$ in (18) of 10^4 randomly selected testing datasets from each agent's database, and obtain each hospital i's MSE in Figure 7 as the mean of $MSE_i(\hat{\theta})$ in (19) of 10^4 randomly selected testing datasets from his database.

B PROOF OF LEMMA 3.1

The min-problem in (3) is equal to

$$\min_{\hat{\theta}} \hat{\theta}^2 - 2\mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\sum_{i=1}^N w_i\theta_i\right]\hat{\theta} + \mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\left(\sum_{i=1}^N w_i\theta_i\right)^2\right],$$

which has an obvious solution of

$$\hat{\theta} = \mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\sum_{i=1}^N w_i\theta_i\right] = \sum_{i=1}^N w_i\mu_i \tag{23}$$

due to its convexity in $\hat{\theta}$. Substituting $\hat{\theta}$ in (23) to (3), we can obtain the system's MSE L_1 in (4) of benchmark 1.

C PROOF OF LEMMA 3.3

Denote $m_{j_1} \le \cdots \le m_{j_N}$ as the agents' increasing, resorted messages $\{m_i\}_{i=1}^N$. At benchmark 2 of the weighted median scheme, the system always determines the global model $\hat{\theta} = m_{j_n}$, where

$$n = \min\left\{k \middle| \sum_{i<k} w_{j_i} \le \frac{1}{2}, \sum_{i>k} w_{j_i} \le \frac{1}{2}, 1 \le k \le N.\right\}.$$

We then have each agent i's loss in (1) of benchmark 2 as follows:

$$\ell_i(\theta_i, m_{j_n}) = \mathbb{E}_{\{\theta_j\}_{j=1,j\neq i}^N, \{m_i(\theta_i)\}_{i=1}^N}\left[(\theta_i - m_{j_n})^2\right], \tag{24}$$

We prove in the following that each agent i cannot obtain a smaller loss by misreporting his parameter θ_i for every possible θ_i and m_{j_n} realization, given the remaining $N-1$ agents truthfully messaging their parameter realizations.

i) Suppose that $\theta_i = m_{j_n}$. In this case, agent i incurs no loss by truthfully reporting his θ_i realization. He thus does not misreport.

ii) Suppose that $\theta_i < m_{j_n}$. In this case, agent i incurs the same loss as truthful messaging if he misreports any $m_i(\theta_i) < \theta_i$ since the weighted median m_{j_n} does not change. Misreporting a $m_i(\theta_i) > \theta_i$ will lead to the weighted median $m'_{j_n} \geq m_{j_n}$, which cannot decrease his loss due to his convex loss and $\theta_i < m_{j_n} \leq m'_{j_n}$. He thus does not misreport.

iii) Suppose that $\theta_i > m_{j_n}$. In this case, agent i incurs the same loss as truthful messaging if he misreports any $m_i(\theta_i) > \theta_i$ since the weighted median m_{j_n} does not change. Misreporting a $m_i(\theta_i) < \theta_i$ will lead to the weighted median $m'_{j_n} \leq m_{j_n}$, which cannot decrease his loss due to his convex loss and $m'_{j_n} \leq m_{j_n} < \theta_i$. He thus does not misreport.

In summary, each agent i cannot obtain a smaller loss by misreporting for every possible θ_i and m_{j_n} realization. To minimize his loss in (24), he will always truthfully message his θ_i realization and the weighted median scheme is thus truthful.

D PROOF OF LEMMA 3.4

Denote $\hat{\theta}_r$ as the random variable of the weighted median parameter and $\tilde{\theta}_{j_n}$ as the random variable of the n_{th} smallest parameter among $\{\theta_i\}_{i=1}^N$. At benchmark 2 of the weighted median scheme, we have the system's global model decision $\hat{\theta} = \hat{\theta}_r$. Substituting this to (2), we obtain the system's MSE of benchmark 2 as follows:

$$\begin{aligned}
L_2 =&\mathbb{E}_{\{\theta_i\}_{i=1}^N,\hat{\theta}_r}\left[\left(\sum_{i=1}^N w_i(\theta_i-\hat{\theta}_r)\right)^2\right] \\
=&\sum_{n=1}^N \mathtt{Pr}(\hat{\theta}_r=\tilde{\theta}_{j_n})\mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\left(\sum_{i=1}^N w_i(\theta_i-\hat{\theta}_r)\right)^2\middle|\hat{\theta}_r=\tilde{\theta}_{j_n}\right] \\
=&\sum_{n=1}^N\sum_{m=1}^N \mathtt{Pr}(\tilde{\theta}_{j_n}=\theta_m)\mathtt{Pr}(\hat{\theta}_r=\tilde{\theta}_{j_n}|\tilde{\theta}_{j_n}=\theta_m)\mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\left(\sum_{i=1}^N w_i(\theta_i-\hat{\theta}_r)\right)^2\middle|\hat{\theta}_r=\theta_m\right] \\
=&\sum_{n=1}^N\sum_{m=1}^N \mathtt{Pr}(\tilde{\theta}_{j_n}=\theta_m)\sum_{m^+\in\Phi_n}\mathtt{Pr}(m^+,\hat{\theta}_r=\tilde{\theta}_{j_n}|\tilde{\theta}_{j_n}=\theta_m)\mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\left(\sum_{i=1}^N w_i(\theta_i-\hat{\theta}_r)\right)^2\middle|\hat{\theta}_r=\theta_m,m^+\right] \\
=&\sum_{n=1}^N\sum_{m=1}^N \mathtt{Pr}(\tilde{\theta}_{j_n}=\theta_m)\sum_{m^+\in\Phi_n}\mathtt{Pr}(m^+|\tilde{\theta}_{j_n}=\theta_m)\mathtt{Pr}(\hat{\theta}_r=\tilde{\theta}_{j_n}|\tilde{\theta}_{j_n}=\theta_m,m^+)\mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\left(\sum_{i\neq m}^N w_i(\theta_i-\theta_m)\right)^2\middle|m^+\right],
\end{aligned} \tag{25}$$

where $m^+ = \{i|\theta_i > \theta_m\}$, $m^- = \{i|\theta_i \leq \theta_m, i \neq m\}$ and Φ_n denotes the set of all m^+s when the weighted median is the n_{th} smallest parameter. Since each agent i's $\theta_i \sim U[A,B]$, we have

$$\mathtt{Pr}(\tilde{\theta}_{j_n}=\theta_m)=\frac{1}{N}, n,m\in[1,N], \tag{26}$$

$$\mathtt{Pr}(m^+|\tilde{\theta}_{j_n}=\theta_m)=\frac{1}{C_{N-1}^{n-1}}, \tag{27}$$

$$\mathtt{Pr}(\hat{\theta}_r=\tilde{\theta}_{j_n}|\tilde{\theta}_{j_n}=\theta_m,m^+)=\mathbf{1}_{\sum_{i\in m^-}w_{j_i}\leq\frac{1}{2}\&\sum_{i\in m^+}w_{j_i}\leq\frac{1}{2}}, \tag{28}$$

and

$$\begin{aligned}
&\mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\left(\sum_{i\neq m}^N w_i(\theta_i-\theta_m)\right)^2\middle|m^+\right] \\
=&\mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\sum_{i\neq m}^N w_i^2(\theta_i-\theta_m)^2+\sum_{i\neq m}^N\sum_{j\neq i,m}^N w_i(\theta_i-\theta_m)w_j(\theta_j-\theta_m)\middle|m^+\right] \\
=&\mathbb{E}_{\{\theta_i\}_{i=1}^N}\left[\sum_{i\in m^+}w_i^2(\theta_i-\theta_m)^2+\sum_{i\in m^-}w_i^2(\theta_i-\theta_m)^2+\left(\sum_{i\neq m^+}w_i(\theta_i-\theta_m)+\sum_{i\neq m^-}w_i(\theta_i-\theta_m)\right)\left(\sum_{j\neq i,j\in m^+}w_j(\theta_j-\theta_m)+\sum_{j\neq i,j\in m^-}w_j(\theta_j-\theta_m)\right)\middle|m^+\right] \\
=&\left(\sum_{i\neq m}^N w_i^2\sigma^2+\frac{3}{5}\sigma^2\left(\sum_{i\neq m,i\in m^+}w_i\sum_{j\neq m,i;j\in m^+}w_j+\sum_{i\neq m,i\in m^-}w_i\sum_{j\neq m,i;j\in m^-}w_j\right)+\frac{1}{10}\sigma^2\left(\sum_{i\neq m,i\in m^+}w_i\sum_{j\neq m,i;j\in m^-}w_j+\sum_{i\neq m,i\in m^-}w_i\sum_{j\neq m,i;j\in m^+}w_j\right)\right).
\end{aligned} \tag{29}$$

Substituting (26)-(29) to (25), we obtain the system's MSE L_2 in (5) as in the lemma. L_2 can be further obtained by introducing $w_i = \frac{1}{N}$ to (5), where we choose the weighted median as $\theta_{\frac{N}{2}}$ for an even N:

$$L_2(N) = \begin{cases} \frac{(N-1)(7N+1)}{20N^2}\sigma^2, & \text{if } N \text{ is odd,} \\ \frac{7N^2-6N+4}{20N^2}\sigma^2, & \text{if } N \text{ is even.} \end{cases} \tag{30}$$

One can easily check that $L_2(N=2) > L_2(N=3)$ and $L_2(N) < L_2(N+1)$ for all $N \geq 3$. We then finish the proof.

E PROOF OF COROLLARY 3.5

Given each agent i's model parameter $\theta_i \sim U[A,B]$ and his identical weight $w_i = \frac{1}{N}$ for $i \in [N]$, we have the system's MSE L_1 in (4) of benchmark 1 as follows:

$$L_1 = \frac{1}{N}\sigma^2. \tag{31}$$

Its MSE L_2 of benchmark 2 is given in (30). One can easily check that L_1 in (31) is larger than L_2 in (30) if $N \in \{2,3\}$, and is smaller if $N \geq 4$. We then finish the proof.

F PROOF OF PROPOSITION 4.3

Let us first simplify each agent i's loss in (1) of our mechanism. The system's global model of our mechanism is $\hat{\theta}(\{\hat{\theta}_i(m_i(\theta_i))\}_{i=1}^N) = \sum_{i=1}^N w_i\hat{\theta}_i(m_i(\theta_i))$. Substituting this to (1), we have

$$\begin{aligned}
&\ell_i(\theta_i, \hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)) \\
=&\mathbb{E}_{\{\theta_j\}_{j=1,j\neq i}^N,\{m_i(\theta_i)\}_{i=1}^N}\left[\left(\theta_i - \hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)\right)^2\right] \\
=&\mathbb{E}_{\{\theta_j\}_{j=1,j\neq i}^N,\{m_i(\theta_i)\}_{i=1}^N}\left[\left(\theta_i - w_i\hat{\theta}_i(m_i(\theta_i)) - \sum_{j\neq i} w_j\hat{\theta}_j(m_j(\theta_j))\right)^2\right] \\
=&\left(\theta_i - w_i\hat{\theta}_i(m_i(\theta_i))\right)^2 + \mathbb{E}_{\{\theta_j\}_{j=1,j\neq i}^N,\{m_i(\theta_i)\}_{i=1}^N}\left[\left(\sum_{j\neq i} w_j\hat{\theta}_j(m_j(\theta_j))\right)^2\right] - 2\left(\theta_i - w_i\hat{\theta}_i(m_i(\theta_i))\right)\sum_{j\neq i} w_j\mathbb{E}_{\theta_j,m_j(\theta_j)}\left[\hat{\theta}_j(m_j(\theta_j))\right].
\end{aligned} \tag{32}$$

Since

$$\begin{aligned}
&\mathbb{E}_{\{\theta_j\}_{j=1,j\neq i}^N,\{m_i(\theta_i)\}_{i=1}^N}\left[\left(\sum_{j\neq i} w_j\hat{\theta}_j(m_j(\theta_j))\right)^2\right] \\
=&\mathbb{E}_{\{\theta_j\}_{j=1,j\neq i}^N,\{m_i(\theta_i)\}_{i=1}^N}\left[\left(\sum_{j\neq i} w_j^2\hat{\theta}_j^2(m_j(\theta_j))\right) + \sum_{j\neq i}\sum_{k\neq j,i} w_jw_k\hat{\theta}_j(m_j(\theta_j))\hat{\theta}_k(m_k(\theta_k))\right] \\
=&\sum_{j\neq i} w_j^2\mathbb{E}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j^2(m_j(\theta_j))] + \sum_{j\neq i}\sum_{k\neq j,i} w_jw_k\mathbb{E}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))]\mathbb{E}_{\theta_k,m_k(\theta_k)}[\hat{\theta}_k(m_k(\theta_k))] \\
=&\sum_{j\neq i} w_j^2\left(\mathbb{E}^2_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))] + \mathbb{V}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))]\right) + \sum_{j\neq i}\sum_{k\neq j,i} w_jw_k\mathbb{E}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))]\mathbb{E}_{\theta_k,m_k(\theta_k)}[\hat{\theta}_k(m_k(\theta_k))] \\
=&\sum_{j\neq i}\left(w_j\mathbb{E}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))]\right)^2 + \sum_{j\neq i}\sum_{k\neq j,i} w_jw_k\mathbb{E}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))]\mathbb{E}_{\theta_k,m_k(\theta_k)}[\hat{\theta}_k(m_k(\theta_k))] + \sum_{j\neq i} w_j^2\left(\mathbb{V}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))]\right) \\
=&\left(\sum_{j\neq i} w_j\mathbb{E}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))]\right)^2 + \sum_{j\neq i} w_j^2\left(\mathbb{V}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))]\right).
\end{aligned} \tag{33}$$

Substitute (33) to (32), we have

$$\begin{aligned}
&\ell_i(\theta_i, \hat{\theta}(\{m_i(\theta_i)\}_{i=1}^N)) \\
=&\left(\theta_i - w_i\hat{\theta}_i(m_i(\theta_i))\right)^2 + \left(\sum_{j\neq i} w_j\mathbb{E}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))\right)^2 - 2\left(\theta_i - w_i\hat{\theta}_i(m_i(\theta_i))\right)\sum_{j\neq i} w_j\mathbb{E}_{\theta_j,m_j(\theta_j)}\left[\hat{\theta}_j(m_j(\theta_j))\right] + \sum_{j\neq i} w_j^2\left(\mathbb{V}_{\theta_j,m_j(\theta_j)}[\hat{\theta}_j(m_j(\theta_j))]\right) \\
=&\left(\theta_i - w_i\hat{\theta}_i(m_i(\theta_i)) - \sum_{j\neq i}^N w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\right)^2 + \sum_{j\neq i}^N w_j\mathbb{V}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))].
\end{aligned} \tag{34}$$

In the following, we first prove the truthfulness of another mechanism, and will then show that it is equivalent to the mechanism in the proposition.

Lemma F.1. *By choosing* $\theta_{i,0} = A_i$, $\theta_{i,K_i} = B_i$ *and* $\{\theta_{i,k}\}_{k=1}^{K_i-1}$ *for any* $K_i \geq 2$ *as the solutions to*

$$\begin{aligned}&\left(\theta_{i,k} - w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]] - \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\right)^2\\ =&\left(\theta_{i,k} - w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k}, \theta_{i,k+1}]] - \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\right)^2,\end{aligned} \tag{35}$$

the system's partial information disclosure mechanism in Definition 4.2 is truthful. Each agent i truthfully messages the index of the partition containing his local model parameter realization, i.e.,

$$m_i^*(\theta_i) = \{k|\theta_i \in [\theta_{i,k-1}, \theta_{i,k}], 1 \leq k \leq K_i.\}$$

Proof. We simplify (35) for further analysis. Since $\{\theta_{i,k}\}_{k=0}^{K_i}$ is a strictly increasing sequence in k, we have

$$\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]] \leq \mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k}, \theta_{i,k+1}]]. \tag{36}$$

We simplify (35) due to (36) as

$$\begin{aligned}&\theta_{i,k} - w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]] - \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\\ =&-\theta_{i,k} + w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k}, \theta_{i,k+1}]] + \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))] > 0.\end{aligned} \tag{37}$$

In the following, we first show that each agent i will not deviate from messaging $m_i(\theta_i) = k$ to $m_i(\theta_i) = k + 1$ if his $\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]$. Since $\theta_i \leq \theta_{i,k}$, according to (37), we have

$$\begin{aligned}&\theta_i - w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]] - \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\\ \leq&-\theta_i + w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k}, \theta_{i,k+1}]] + \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))].\end{aligned} \tag{38}$$

His loss in (34) with his message $m_i(\theta_i) = k$ is

$$\begin{aligned}&\ell_i(\theta_i|m_i(\theta_i) = k)\\ =&\left(\theta_i - w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]]) - \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\right)^2\\ &+\sum_{j\neq i}^{N} w_j\mathbb{V}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\end{aligned} \tag{39}$$

His loss in (34) with his message $m_i(\theta_i) = k + 1$ is

$$\begin{aligned}&\ell_i(\theta_i|m_i(\theta_i) = k+1)\\ =&\left(\theta_i - w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k}, \theta_{i,k+1}]]) - \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\right)^2\\ &+\sum_{j\neq i}^{N} w_j\mathbb{V}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\end{aligned} \tag{40}$$

To show $\ell_i(\theta_i|m_i(\theta_i) = k) \leq \ell_i(\theta_i|m_i(\theta_i) = k + 1)$, according to (39) and (40), we only need to show that

$$\begin{aligned}&\left(\theta_i - w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k-1}, \theta_{i,k}]]) - \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\right)^2\\ \leq&\left(\theta_i - w_i\mathbb{E}[\theta_i|\theta_i \in [\theta_{i,k}, \theta_{i,k+1}]]) - \sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\right)^2,\end{aligned}$$

which holds due to (38). We further have $\ell_i(\theta_i|m_i(\theta_i)=k)\leq \ell_i(\theta_i|m_i(\theta_i)=k')$ for any $k'\geq k+1$ due to

$$\begin{aligned}&\theta_i - w_i\mathbb{E}[\theta_i|\theta_i\in[\theta_{i,k-1},\theta_{i,k}]]-\sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\\ \leq & -\theta_i + w_i\mathbb{E}[\theta_i|\theta_i\in[\theta_{i,k},\theta_{i,k+1}]]+\sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\\ \leq & -\theta_i + w_i\mathbb{E}[\theta_i|\theta_i\in[\theta_{i,k'-1},\theta_{i,k'}]]+\sum_{j\neq i}^{N} w_j\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))].\end{aligned}$$

With a similar analysis as above, we can obtain that $\ell_i(\theta_i|m_i(\theta_i)=k)\leq \ell_i(\theta_i|m_i(\theta_i)=k-1)$ and $\ell_i(\theta_i|m_i(\theta_i)=k)\leq \ell_i(\theta_i|m_i(\theta_i)=k')$ for any $k'\leq k-1$. Therefore, each agent i incurs the smallest loss by truthful messaging:

$$m_i^*(\theta_i)=\{k|\theta_i\in[\theta_{i,k-1},\theta_{i,k}],1\leq k\leq K_i.\}$$

We then finish the proof. □

According to Lemma F.1, each agent i truthfully messages the partition containing his θ_i realization with the mechanism in (35). Therefore, we have

$$\begin{aligned}&\mathbb{E}_{\theta_j,m_j|\theta_j}[\hat{\theta}_j(m_j(\theta_j))]\\ =&\sum_{k=1}^{K_i}\mathtt{Pr}(\theta_j\in[\theta_{j,k-1},\theta_{j,k}])\mathbb{E}[\theta_j|\theta_j\in[\theta_{j,k-1},\theta_{j,k}]]=\mathbb{E}[\theta_j]=\mu_j.\end{aligned}\tag{41}$$

Substituting (41) to (35), we can obtain (6) in the proposition, which implies that our mechanism in (6) is truthful. We then finish the proof.

G PROOF OF PROPOSITION 4.4

Note that (8) is a linear, autonomous, second-order difference equation. We divide the equations in (8) into two parts: steady-state part and homogeneous part, and the solution to (8) is the sum of solutions of the two parts (e.g., [16]). The steady-state part implies that $\theta_{i,t}=\bar{\theta}_i$ for all $t\geq 0$. We substitute it to (8) and obtain

$$\bar{\theta}_i=\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}.\tag{42}$$

The homogeneous part of the difference equation is given as follows:

$$\theta_{i,k+1}-\frac{2(2-w_i)}{w_i}\theta_{i,k}+\theta_{i,k-1}=0.\tag{43}$$

(43) has a trivial solution of $\theta_{i,k}=0$. The non-trivial homogeneous solution has form of $\theta_{i,k}=c\lambda^k$, where $c\neq 0$ and $\lambda\neq 0$ are two constants (e.g., [16]). We substitute $\theta_{i,k}$ to (43) and obtain

$$\lambda^2-\frac{2(2-w_i)}{w_i}\lambda+1=0,\tag{44}$$

which is known as the characteristic equation of the homogeneous difference equation. The solutions to (44) are known as the characteristic roots, which are

$$\lambda_1=\frac{(1+\sqrt{1-w_i})^2}{w_i},\ \lambda_2=\frac{(1-\sqrt{1-w_i})^2}{w_i}.\tag{45}$$

Note that the two solutions in (45) are linearly independent. Therefore, the homogeneous solution $\theta_{i,h}$ to (43) has the following form:

$$\theta_{i,h}=\alpha_i\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^k+\beta_i\left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^k,\tag{46}$$

where α_i and β_i are constants. Together with (42), we are now able to give the solution to (8) as follows:

$$\theta_{i,k}=\alpha_i\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^k+\beta_i\left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^k+\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}.\tag{47}$$

We can solve α_i and β_i in (47) according to the boundary conditions $\theta_{i,0}=A_i$ and $\theta_{i,K_i}=B_i$. We then finish the proof.

H PROOF OF COROLLARY 4.5

The system's MSE $L^*(\{K_i\}_{i=1}^N)$ in (10) can be obtained according to (2) and Proposition 4.4. Note that $L^*(\{K_i\}_{i=1}^N)$ is separable in each K_i. We thus define its MSE $\bar{L}_i^*(K_i)$ from each agent i as follows:

$$\bar{L}_i^*(K_i) = \frac{w_i(1-w_i)}{3(4-w_i)\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)^2}\left(24\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - \mu_i\right)^2 + 6(A_i-B_i)^2 - 12\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - A_i\right)\right.$$
$$\left.\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - B_i\right)\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)\right). \tag{48}$$

Note that we ignore the last term in (10) because it is independent of K_i. In the following, we prove that $L^*(\{K_i\}_{i=1}^N)$ decreases with each K_i case by case.

i) If $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in [A_i, B_i]$, we have the numerator of $\bar{L}_i^*$ in (48) is always positive. Note that $\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}$ increases with K_i due to $w_i \in (0,1)$, $\frac{(1+\sqrt{1-w_i})^2}{w_i} \in (1,\infty)$ and $\frac{(1-\sqrt{1-w_i})^2}{w_i} \in (0,1)$. Accordingly, we have $\bar{L}_i^*$ in (48) decreases with K_i if $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in [A_i, B_i]$.

ii) If $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in (-\infty, A_i)\bigcup(B_i,\infty)$, the derivative of $\bar{L}_i^*$ in (48) is

$$\frac{\partial \bar{L}_i^*}{\partial K_i} = \frac{-12w_i(1-w_i)\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - A_i\right)\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - B_i\right)}{3(4-w_i)\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)^2}\frac{\partial\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)}{\partial K_i}$$
$$+\frac{w_i(1-w_i)}{3(4-w_i)}\frac{\partial\frac{1}{\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)^2}}{\partial K_i}$$
$$\left(24\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - \mu_i\right)^2 + 6(A_i-B_i)^2 - 12\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - A_i\right)\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - B_i\right)\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)\right).$$

Note that the first term of the above $\frac{\partial \bar{L}_i^*}{\partial K_i}$ is negative due to $w_i \in (0,1)$, $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in (-\infty, A_i)\bigcup(B_i,\infty)$ and $\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}$ increasing in K_i. We can thus show $\frac{\partial \bar{L}_i^*}{\partial K_i} < 0$ by showing its second term is negative. Since $\frac{\partial\frac{1}{\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)^2}}{\partial K_i} < 0$ due to $\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}$ increasing in K_i, we only need to show that

$$\left(24\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - \mu_i\right)^2 + 6(A_i-B_i)^2 - 12\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - A_i\right)\right.$$
$$\left.\left(\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} - B_i\right)\left(\left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} - \left(\frac{(1-\sqrt{1-w_i})^2}{w_i}\right)^{K_i}\right)\right) > 0,$$

which is equivalent to

$$\frac{B_i - \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}}{A_i - \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}}x^2 + \left(\left(\frac{B_i - \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}}{A_i - \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}}\right)^2 + 1\right)x - \frac{B_i - \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}}{A_i - \frac{\sum_{j\neq i} w_j\mu_j}{1-w_i}} > 0, \tag{49}$$

where

$$x = \left(\frac{(1+\sqrt{1-w_i})^2}{w_i}\right)^{K_i} > 1, w_i \in (0,1), K_i \geq 1.$$

Note that the left-handed side of (49) increases with $x > 1$. Further, it is positive when $x = 1$. Therefore, we have (49) holds for any $K_i \geq 1$, which implies $\bar{L}_i^*$ in (48) decreases with K_i if $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in (-\infty, A_i)\bigcup(B_i,\infty)$.

In summary, we have $\bar{L}_i^*$ in (48) decreases with K_i and L^* in (10) thus decreases with each K_i.

I PROOF OF PROPOSITION 4.6

We first find the feasible range of K_i to satisfy that $\{\theta_{i,k}\}_{k=0}^{K_i}$ is strictly increasing in k, which is equivalent to find the condition for $\theta_{i,k+1} - \theta_{i,k} > 0$ for all $k \in [0, K_i - 1]$. We have

$$\begin{aligned}\theta_{i,k+1} - \theta_{i,k} = &\frac{x-1}{x(x^{-K_i} - x^{K_i})}\Big(x^{1+k}(x^{-K_i}(A_i - \bar{\mu}_{-i}) - (B_i - \bar{\mu}_{-i}))\\ &+ x^{-k}(x^{K_i}(A_i - \bar{\mu}_{-i}) - (B_i - \bar{\mu}_{-i}))\Big),\end{aligned} \tag{50}$$

where $x = \frac{(1+\sqrt{1-w_i})^2}{w_i} > 1$ for $w_i \in (0,1)$. Note that $\frac{x-1}{x(x^{-K_i}-x^{K_i})} < 0$ for $x > 1$ and $K_i \geq 1$. To find the feasible range of K_i such that $\theta_{i,k+1} - \theta_{i,k} > 0$, it is equivalent to find the condition of K_i to satisfy

$$\begin{aligned}g(k) :=&x^{1+k}(x^{-K_i}(A_i - \bar{\mu}_{-i}) - (B_i - \bar{\mu}_{-i}))\\ &+ x^{-k}(x^{K_i}(A_i - \bar{\mu}_{-i}) - (B_i - \bar{\mu}_{-i})) < 0.\end{aligned} \tag{51}$$

i) If $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in (A_i, B_i)$, note that $g(k)$ in (51) is always negative with any $k \geq 0$. Therefore, the feasible range of K_i is $[1, \infty]$.

ii) If $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in (-\infty, A_i]$, we have $x^{-K_i}(A_i-\bar{\mu}_{-i})-(B_i-\bar{\mu}_{-i}) < 0$ due to $x > 1$. $g(k)$ in (51) is always negative if $x^{K_i}(A_i-\bar{\mu}_{-i})-(B_i-\bar{\mu}_{-i}) < 0$, i.e., $K_i \leq \log_x \frac{B_i-\bar{\mu}_{-i}}{A_i-\bar{\mu}_{-i}}$. If $K_i > \log_x \frac{B_i-\bar{\mu}_{-i}}{A_i-\bar{\mu}_{-i}}$, we have $g(k)$ decreasing in k due to $g'(k) < 0$. To ensure $g(k) < 0$ for all $k \geq 0$, we need to satisfy

$$g(k=0) = (A_i - \bar{\mu}_{-i})\left(\frac{x}{x^{K_i}} + x^{K_i} - (x+1)\frac{B_i - \bar{\mu}_{-i}}{A_i - \bar{\mu}_{-i}}\right) < 0,$$

which is equivalent to

$$K_i \leq \log_x \frac{(x+1)\frac{B_i-\bar{\mu}_{-i}}{A_i-\bar{\mu}_{-i}} + \sqrt{\left((x+1)\frac{B_i-\bar{\mu}_{-i}}{A_i-\bar{\mu}_{-i}}\right)^2 - 4x}}{2}. \tag{52}$$

One can easily show the following inequality holds:

$$\frac{(x+1)\frac{B_i-\bar{\mu}_{-i}}{A_i-\bar{\mu}_{-i}} + \sqrt{\left((x+1)\frac{B_i-\bar{\mu}_{-i}}{A_i-\bar{\mu}_{-i}}\right)^2 - 4x}}{2} > \frac{B_i - \bar{\mu}_{-i}}{A_i - \bar{\mu}_{-i}}, x > 1.$$

Therefore, the feasible range of K_i for $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in (-\infty, A_i]$ is (52).

iii) If $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in [B_i, \infty)$, we have $x^{K_i}(A_i-\bar{\mu}_{-i})-(B_i-\bar{\mu}_{-i}) < 0$ due to $x > 1$. $g(k)$ in (51) is always negative if $x^{-K_i}(A_i-\bar{\mu}_{-i})-(B_i-\bar{\mu}_{-i}) < 0$, i.e., $K_i \leq \log_x \frac{A_i-\bar{\mu}_{-i}}{B_i-\bar{\mu}_{-i}}$. If $K_i > \log_x \frac{A_i-\bar{\mu}_{-i}}{B_i-\bar{\mu}_{-i}}$, we have $g(k)$ increasing in k due to $g'(k) > 0$. To ensure $g(k) < 0$ for all $k \geq 0$, we need to satisfy

$$g(k = K_i - 1) = (-A_i + \bar{\mu}_{-i})\left((\frac{x}{x^{K_i}} + x^{K_i})\frac{B_i - \bar{\mu}_{-i}}{A_i - \bar{\mu}_{-i}} - (x+1)\right) < 0,$$

which is equivalent to

$$K_i \leq \log_x \frac{(x+1)\frac{A_i-\bar{\mu}_{-i}}{B_i-\bar{\mu}_{-i}} + \sqrt{\left((x+1)\frac{A_i-\bar{\mu}_{-i}}{B_i-\bar{\mu}_{-i}}\right)^2 - 4x}}{2}. \tag{53}$$

One can easily show the following inequality holds:

$$\frac{(x+1)\frac{A_i-\bar{\mu}_{-i}}{B_i-\bar{\mu}_{-i}} + \sqrt{\left((x+1)\frac{A_i-\bar{\mu}_{-i}}{B_i-\bar{\mu}_{-i}}\right)^2 - 4x}}{2} > \frac{A_i - \bar{\mu}_{-i}}{B_i - \bar{\mu}_{-i}}, x > 1.$$

Therefore, the feasible range of K_i for $\frac{\sum_{j\neq i} w_j\mu_j}{1-w_i} \in (B_i, \infty)$ is (53).

According to Corollary 4.5, we have the system's MSE decreases with each K_i. Therefore, it is optimal for its to determine each integer K_i as large as possible in the feasible range. We then have K_i^* in (11) and finish the proof.

J PROOF OF COROLLARY 4.7

Corollary 4.7 can be obtained directly from K_i^* in (11). We skip the trivial proof here.

K PROOF OF PROPOSITION 4.8

Proposition 4.8 can be obtained directly from Propositions 4.4 and 4.6. We skip the trivial proof here.

L PROOF OF LEMMA 5.1

The system's MSE L^* in (13) can be obtained according to Corollary 4.5 and Proposition 4.6. If $w_i = \frac{1}{N}$ and $\sigma_i^2 = \sigma^2$ for $i \in [N]$, L^* in (13) becomes

$$L^*(N) = \frac{N-1}{(4N-1)N}\sigma^2,$$

which decreases with $N \geq 2$ due to $\frac{\partial L^*(N)}{\partial N} < 0$. It is obvious to see that $\lim_{N\to\infty} L^*(N) = 0$. We then finish the proof.

M PROOF OF PROPOSITION 5.2

Note that $\frac{1-w_i}{4-w_i}$ decreases with $w_i \in (0,1)$, which implies

$$\frac{1-w_i}{4-w_i} < \frac{1}{4}, w_i \in (0,1).$$

Therefore, we have r_1 in (14) as follows:

$$r_1 = \frac{\sum_{i=1}^N \frac{(1-w_i)}{4-w_i} w_i^2\sigma_i^2}{\sum_{i=1}^N w_i^2\sigma_i^2} < \frac{\sum_{i=1}^N \frac{1}{4} w_i^2\sigma_i^2}{\sum_{i=1}^N w_i^2\sigma_i^2} = \frac{1}{4}.$$

If $\theta_i \in U[A,B]$ and $w_i = \frac{1}{N}$ for $i \in [N]$, r_1 in (14) becomes

$$r_1(N) = \frac{N-1}{4N-1},$$

which increases with N due to $r_1'(N) > 0$. It is obvious to see that $\lim_{N\to\infty} r_1(N) = \frac{1}{4}$.

N PROOF OF PROPOSITION 5.3

r_2 in (15) can be directly obtained according (13) and (5). It is easy to check that $r_2(N-1) > r_2(N)$ for any $N \geq 2$. It is also obvious to see that $\lim_{N\to\infty} r_2(N) = 0$. We skip the trivial proof here.

O PROOF OF LEMMA 5.4

If $K_i = 2$, (6) becomes:

$$2\theta_{i,1} - w_i \frac{\int_{x=A_i}^{\theta_{i,1}} x f_i(x)dx}{F_i(\theta_{i,1})} - w_i \frac{\int_{x=\theta_{i,1}}^{B_i} x f_i(x)dx}{1-F_i(\theta_{i,1})} = 2\sum_{j\neq i}^{N} w_j\mu_j. \tag{54}$$

Define a function $g(\theta_{i,1})$ to represent the left-handed side of (54):

$$g(\theta_{i,1}) = 2\theta_{i,1} - w_i \frac{\int_{x=A_i}^{\theta_{i,1}} x f_i(x)dx}{F_i(\theta_{i,1})} - w_i \frac{\int_{x=\theta_{i,1}}^{B_i} x f_i(x)dx}{1-F_i(\theta_{i,1})}.$$

Since the PDF function $f_i(\cdot)$ is continuous as defined, we have the CDF function $F_i(\cdot)$ is also continuous. Further, the integrals $\int_{x=A_i}^{\theta_{i,1}} x f_i(x)dx$ and $\int_{x=\theta_{i,1}}^{B_i} x f_i(x)dx$ are continuous, too. According to the L'Hôpital's rule, we have

$$\lim_{\theta_{i,1}\to A_i} \frac{\int_{x=A_i}^{\theta_{i,1}} x f_i(x)dx}{F_i(\theta_{i,1})} = \lim_{\theta_{i,1}\to A_i} \frac{\frac{\partial \int_{x=A_i}^{\theta_{i,1}} x f_i(x)dx}{\partial\theta_{i,1}}}{\frac{\partial F_i(\theta_{i,1})}{\partial\theta_{i,1}}} = A_i,$$

$$\lim_{\theta_{i,1}\to B_i} \frac{\int_{x=\theta_{i,1}}^{B_i} x f_i(x)dx}{1-F_i(\theta_{i,1})} = \lim_{\theta_{i,1}\to B_i} \frac{\frac{\partial \int_{x=\theta_{i,1}}^{B_i} x f_i(x)dx}{\partial\theta_{i,1}}}{\frac{\partial(1-F_i(\theta_{i,1}))}{\partial\theta_{i,1}}} = B_i.$$

Therefore, $g(\cdot)$ is a continuous function. Besides, we have

$$g(A_i) = (2-w_i)A_i - w_i\mu_i < 2\sum_{j\neq i}^{N} w_j\mu_j,$$

$$g(B_i) = (2-w_i)B_i - w_i\mu_i > 2\sum_{j\neq i}^{N} w_j\mu_j,$$

from (54). According to the Intermediate value theorem, there exists at least one $\theta_{i,1}^*$ satisfying $g(\theta_{i,1}^*) = 2\sum_{j\neq i}^N w_j\mu_j$, which proves the existence of $\theta_{i,1}$ for $K_i = 2$. Therefore, the largest feasible K_i is no less than 2. We then finish the proof.

P PROOF OF LEMMA 5.5

Lemma 5.5 is obtained according to the following theorem.

Theorem P.1 (Theorem 3.5 in Section 3.3 of [22]). *Given $\mathbf{h}_i : \mathcal{R}^{K_i-1} \to \mathcal{R}^{K_i-1}$ is continuously differentiable and $\boldsymbol{\theta}_i^*$ is a root of $\mathbf{h}_i$ (i.e., $\mathbf{h}_i(\boldsymbol{\theta}_i^*) = 0$) such that its Jacobian matrix $\mathbf{J}_i(\boldsymbol{\theta}_i^*)$ is non-singular. Then there exists a positive $\delta > 0$ such that if $||\mathbf{s}_i^0 - \boldsymbol{\theta}_i^*|| < \delta$, we have*

$$\lim_{t\to\infty} \frac{||\mathbf{s}_i^{t+1} - \boldsymbol{\theta}_i^*||}{||\mathbf{s}_i^t - \boldsymbol{\theta}_i^*||} = 0.$$

If $\mathbf{J}_i(\cdot)$ is further Lipschitz continuous, there exists a positive $M > 0$ such that for all $t \geq 0$,

$$||\mathbf{s}_i^{t+1} - \boldsymbol{\theta}_i^*|| \leq M||\mathbf{s}_i^t - \boldsymbol{\theta}_i^*||^2.$$

The Jacobian matrix $\mathbf{J}_i(\boldsymbol{\theta}_i^*)$ under a truncated normal distribution is non-singular due to the rank of this matrix is full, which is easy to check. Besides, the Jacobian matrix $\mathbf{J}_i(\boldsymbol{\theta}_i^*)$ under a truncated normal distribution is Lipschitz continuous since each $\mathbf{J}_i(\cdot)$ is differentiable and has bounded first derivative. We then finish the proof.

Q PROOF OF PROPOSITION 6.2

Proposition 6.2 can be proved by following the same steps as in the proof of Proposition 4.3 in Appendix F. We skip the redundancy here.

R PROOF OF PROPOSITION 6.3

It is enough to prove that $\mathbb{E}_{\{\theta_i\}_{i=1}^N,\{n_i\}_{i=1}^N}[L^*]$ in ΔL^* in (16) has a order of $O(D^2)$ to prove that ΔL^* has such order, since the second term in ΔL^* is independent of D. Suppose that $\theta_i \in U[\hat{A}_i, \hat{B}_i]$ and $A_i + D \in [\theta_{i,M}, \theta_{i,M+1}]$, where $\hat{A}_i \in [A_i - D, A_i + D]$, $M \leq \lfloor N/2 \rfloor$ and $\theta_{i,M}$ is given in Proposition 6.2. Accordingly, We have

$$\begin{aligned}
&\mathbb{E}_{\{\theta_i\}_{i=1}^N,\{n_i\}_{i=1}^N}[L^*] \\
=&\sum_{i=1}^N \Bigg(\sum_{m=0}^{M-1} \int_{\hat{A}_i=\theta_{i,m}}^{\theta_{i,m+1}} \frac{1}{2D(A_i+B_i-2\hat{A}_i)} \left(\sum_{k=m}^{K_i-m-2} \frac{2}{3}\left(\frac{\theta_{i,k+1}-\theta_{i,k}}{2}\right)^3 + \frac{2}{3}\left(\frac{\theta_{i,m}+\theta_{i,m+1}}{2} - \hat{\theta}_i\right)^3 \right) d\hat{\theta}_i \\
&\quad + \int_{\hat{A}_i=\theta_{i,M}}^{A_i+D} \frac{1}{2D(A_i+B_i-2\hat{A}_i)} \left(\sum_{k=M}^{K_i-M-2} \frac{2}{3}\left(\frac{\theta_{i,k+1}-\theta_{i,k}}{2}\right)^3 + \frac{2}{3}\left(\frac{\theta_{i,M}+\theta_{i,M+1}}{2} - \hat{\theta}_i\right)^3 \right) d\hat{\theta}_i \Bigg) \\
<&\sum_{i=1}^N \Bigg(\sum_{m=0}^{M-1} \int_{\hat{A}_i=\theta_{i,m}}^{\theta_{i,m+1}} \frac{1}{2D(A_i+B_i-2(A_i+D))} \left(\sum_{k=m}^{K_i-m-2} \frac{2}{3}\left(\frac{\theta_{i,k+1}-\theta_{i,k}}{2}\right)^3 + \frac{2}{3}\left(\frac{\theta_{i,m}+\theta_{i,m+1}}{2} - \hat{\theta}_i\right)^3 \right) d\hat{\theta}_i \\
&\quad + \int_{\hat{A}_i=\theta_{i,M}}^{A_i+D} \frac{1}{2D(A_i+B_i-2(A_i+D))} \left(\sum_{k=M}^{K_i-M-2} \frac{2}{3}\left(\frac{\theta_{i,k+1}-\theta_{i,k}}{2}\right)^3 + \frac{2}{3}\left(\frac{\theta_{i,M}+\theta_{i,M+1}}{2} - \hat{\theta}_i\right)^3 \right) d\hat{\theta}_i \Bigg) \\
=&\frac{1}{2D(A_i+B_i-2(A_i+D))} \left(\sum_{m=0}^{M-1} \sum_{k=m}^{K_i-m-2} \frac{2}{3}\left(\frac{\theta_{i,k+1}-\theta_{i,k}}{2}\right)^3 (\theta_{i,m+1}-\theta_{i,m}) \right. \\
+&\left. \sum_{k=M}^{K_i-M-2} \frac{2}{3}\left(\frac{\theta_{i,k+1}-\theta_{i,k}}{2}\right)^3 (A_i+D-\theta_{i,M}) - \frac{1}{6}\left(\left(\frac{\theta_{i,M}+\theta_{i,M+1}}{2} - (A_i+D)\right)^4 - \left(\frac{\theta_{i,M}+\theta_{i,M+1}}{2}\right)^4 \right) \right).
\end{aligned}$$

According to (12), we have each $\theta_{i,k}$ is linear in D under uniform distribution and identical mean $\mu_i = \mu$. Besides, the numerator of the above $\mathbb{E}_{\{\theta_i\}_{i=1}^N,\{n_i\}_{i=1}^N}[L^*]$ has the order of $O(D^4)$ and the denominator has the order of D^2, which implies that $\mathbb{E}_{\{\theta_i\}_{i=1}^N,\{n_i\}_{i=1}^N}[L^*]$ incurs a complexity order of $O(D^2)$. ΔL^* in (16) thus has a order of $O(D^2)$.