Table 6: Overall Reputation with # workers/ #requester = 5

	Dadaset	#Req.	BFL(CO/NC)	RRAFL	RanPricing
Coop.	MINIST	4	**10.30**	4.81	0.62
		6	**13.12**	5.62	0.91
		8	**17.15**	6.32	0.83
		10	**18.91**	7.14	1.01
	Fashion MINIST	4	**9.12**	4.31	0.61
		6	**14.81**	7.22	0.83
		8	**16.53**	8.32	0.92
		10	**18.13**	9.51	1.02
Non-coop.	MINIST	4	**9.45**	4.34	0.51
		6	**11.12**	5.12	0.72
		8	**15.08**	6.11	0.83
		10	**17.09**	6.89	0.97
	Fashion MINIST	4	**8.26**	3.92	0.58
		6	**11.12**	4.15	0.73
		8	**13.25**	5.20	0.42
		10	**16.21**	6.31	1.03

Table 7: Average Global Accuracy with # workers/ #requester = 5

	Dadaset	#Req.	BFL(CO/NC)	RRAFL	RanPricing
Coop.	MINIST	4	**0.862**	0.735	0.553
		6	**0.841**	0.734	0.462
		8	**0.849**	0.627	0.428
		10	**0.832**	0.709	0.425
	Fashion MINIST	4	**0.747**	0.591	0.443
		6	**0.739**	0.625	0.491
		8	**0.716**	0.63	0.459
		10	**0.728**	0.615	0.417
Non-coop.	MINIST	4	**0.848**	0.652	0.484
		6	**0.822**	0.612	0.474
		8	**0.841**	0.575	0.42
		10	**0.827**	0.583	0.414
	Fashion MINIST	4	**0.735**	0.549	0.437
		6	**0.711**	0.518	0.474
		8	**0.708**	0.512	0.415
		10	**0.723**	0.506	0.383

B REBUTTAL APPENDIX

B.1 Novelty and Technique Quality

Novelty: In FL, it has been observed that workers are not willing to contribute their raw data unconditionally for training the local model due to the costs associated with data collection and computational resource consumption. Therefore, it is crucial to design incentive mechanisms in FL that provide monetary rewards as compensation to incentivize worker participation. The challenge in designing incentive mechanisms is that workers' costs are private and the designed mechanism should guarantee truthfulness (i.e., ensuring that workers always report their true costs). Additionally, an important aspect is efficiently selecting high-quality workers to complete the training task. However, existing literature only focuses on the incentive mechanism in FL for a single requester and ignores compatibility constraints among workers, such as conflicts in communication channels or conflicting interests. These compatibility constraints can significantly impact the accuracy of the trained global model. Thus, this paper aims to address these limitations by proposing budget-constrained incentive mechanisms that consider multiple requesters with budgets and the heterogeneity of workers in real-world FL systems. Our proposed mechanisms aim to improve the efficiency of selecting high-quality workers while considering budget constraints and compatibility issues.

Technique: From the technique aspect, as requesters can cooperate with each other to maximize the overall performance or be selfish to maximize their own utilities [35], we consider two different settings depending on requesters' behavior: (i) The cooperative budget setting where requesters cooperatively share their own budgets and ii) The non-cooperative budget setting where each requester is unwilling to share the budget and wants to hire workers under their own budget. Considering the above settings, our main technical contributions are listed as follows:

For (i), we propose a mechanism that transforms the allocation of workers within compatibility constraints into a max-flow problem. This allows us to explore different potential prices while simultaneously ensuring efficiency, budget feasibility, and truthfulness. In addition, the proposed mechanism achieves a constant approximation ratio (compared to the optimal sum of reputation obtained by the optima solution).

For (ii), in this context, we utilize the concept of virtual prices to evaluate requesters' procurement ability and propose mechanisms for determining the critical price that aligns with their procurement ability. It is non-trivial to choose the appropriate critical price that can ensure both budget feasibility and truthfulness. The proposed mechanisms all ensure approximation guarantees.

We prove that our mechanisms guarantee computational efficiency, individual rationality, budget feasibility, truthfulness, and approximation to the optimal solution with respect to the sum of chosen workers' reputation.

We also conduct experiments on real-world datasets, MNIST and Fashion MNIST, which are commonly used in FL and data mining. The simulation results show that our mechanisms outperform existing benchmarks in terms of the overall reputation of selected workers and the average accuracy of requesters' global models.

B.2 Reviewer aZoM

For scalability: Table 6 and Table 7 show the overall reputation and the accuracy with the increase of requesters under the fixed ratio between the number of workers and requesters. Although there is a slight decrease in average accuracy with an increasing number of requesters, our proposed mechanisms always significantly outperform the baseline mechanisms.

For biased allocation: We propose a simple biased allocation method called GreedyPri, which prioritizes allocating workers to requesters with higher budgets. Table 8 presents the overall reputation and accuracy results. It is evident that the performance of

Table 8: Performance of GreedyPri

	Dadaset	#Req.	Reputation	Accuracy
Coop.	MINIST	4	0.45	0.621
		6	0.74	0.586
		8	0.96	0.588
		10	1.38	0.552
	Fashion MINIST	4	1.08	0.492
		6	1.27	0.443
		8	1.82	0.442
		10	1.91	0.407
Non-coop.	MINIST	4	0.43	0.629
		6	0.62	0.572
		8	0.91	0.535
		10	1.11	0.541
	Fashion MINIST	4	0.88	0.49
		6	1.16	0.445
		8	1.51	0.449
		10	2.17	0.414

GreedyPri is significantly inferior to that of our proposed mechanisms.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009