Table 6: Key Notations

Notation	Description
A	set of requesters
S	set of workers
c_i	the private cost of worker s_i
v_i	the reputation of worker s_i
G_l	the set of workers with type l
g_{max}	the maximum worker size among different types
B_j	the budget of requester a_j
b_i	the bid of worker s_i
$x_{ij} \in \{0, 1\}$	indicate whether s_i is allocated to a_j
p_{ij}	the payment paid to s_i from a_j
$u_s^i(\cdot)$	the utility of worker s_i
ρ_i	the virtual reputation of s_i
D_h	the set of workers with reputation in $(\varepsilon^{h-1}, \varepsilon^h]$
n_h	the size of worker set D_h
ρ_{max}	the maximum virtual repuration

B REBUTTAL APPENDIX

B.1 Relevance to KDD

- Data mining involves the discovery of patterns and knowledge from datasets by considering methods at the intersection of machine learning, statistics, and database systems. In FL, where data is distributed across multiple workers or clients, the selection of high-quality workers to obtain large/good data sets becomes crucial to facilitate building good machine learning models (which is a crucial task within data mining).
- Our research contributes and is very relevant to data mining because it provides effective strategies to select workers and incentivizes them to provide datasets in FL, ensuring the inclusion of high-quality datasets in the (supervised) learning process.
- Furthermore, the application of FL extends beyond traditional centralized data mining approaches to decentralized ones. Our research aligns with this paradigm, providing insights and methodologies that can enhance the efficiency and effectiveness of data mining in decentralized settings. Moreover, we also conduct experiments to show the strong performance of our mechanisms for several classification problems on real datasets, MNIST and Fashion MNIST, which are commonly used in FL and data mining.

In addition to its relevance to data mining, our paper also contributes to a growing list of literature at the intersection of data mining and interdisciplinary areas (i.e., budget feasible mechanisms, incentive mechanism design, and procurement markets). The literature includes publications in data mining conferences or journals such as KDD, ICDM, WSDM, and TKDE (see below the list of selected papers). Thus, our work provides valuable insights and an understanding of the interplay between data mining and these areas.

Budget feasible (or budget-constrained incentive) mechanisms

- Singer Y. How to win friends and influence people, truthfully: influence maximization mechanisms for social networks, WSDM

Incentive mechanisms

- Y. Lu, Q. Tang, and G. Wang, On Enabling Machine Learning Tasks atop Public Blockchains: A Crowdsourcing Approach, ICDM
- Zhang M, Li X, Miao Y, et al. Peak: Privacy-enhanced incentive mechanism for distributed k-anonymity in LBS, TKDE
- Toyoda K, Zhang A N. Mechanism design for an incentive-aware blockchain-enabled federated learning platform, Big Data
- Quan S, Tan H, Liu S, et al. MERIT: A Merchant Incentive Ranking Model for Hotel Search Ranking, CIKM
- Muldoon C, O'Grady M J, O'Hare G M P. A survey of incentive engineering for crowdsourcing, The Knowledge Engineering Review

Procurement (or budget-constrained procurement) markets

- Anagnostopoulos A, Castillo C, Fazzone A, et al. Algorithms for hiring and outsourcing in the online labor market, KDD

Mechanism design

- Zhu T, Li J, Hu X, et al. The dynamic privacy-preserving mechanisms for online dynamic social networks, TKDE
- Grubenmann T, Cheng R C K, Lakshmanan L V S. TSA: A truthful mechanism for social advertising, WSDM
- Jung K, Lee J, Park K, et al. PRIVATA: differentially private data market framework using negotiation-based pricing mechanism, CIKM
- Conitzer V. Automated Mechanism Design for Strategic Classification: Abstract for KDD'21 Keynote Talk, KDD
- Li Y, Miao C, Su L, et al. An efficient two-layer mechanism for privacy-preserving truth discovery, KDD

Auction design

- Li N, Ma Y, Zhao Y, et al. Learning-Based Ad Auction Design with Externalities: The Framework and A Matching-Based Approach, KDD
- Liu X, Yu C, Zhang Z, et al. Neural auction: End-to-end learning of auction mechanisms for e-commerce advertising, KDD
- Chen Y, Liu W, Yi J, et al. Query clustering based on bid landscape for sponsored search auction optimization, KDD

Games

- Tao J, Lin J, Zhang S, et al. Mvan: Multi-view attention networks for real money trading detection in online games, KDD
- Brückner M, Scheffer T. Stackelberg games for adversarial prediction problems, KDD

B.2 Novelty

Novelty.

We have two main points of innovation (details are provided below and also provided in the paper). The first point is conceptional through new models capturing realistic FL learning settings. The second point is the design of new incentive mechanisms with

Table 7: Overall Reputation with # workers/ #requester = 5

	Dadaset	#Req.	BFL(CO/NC)	RRAFL	RanPri
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 8: Average Global Accuracy with # workers/ #requester = 5

	Dadaset	#Req.	BFL(CO/NC)	RRAFL	RanPri
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

theoretical and empirical guarantees for facilitating FL more effectively. None of the models and mechanisms has been considered previously.

Conceptual 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 [29, 38, 39]. **Therefore, we consider the design of incentive mechanisms in FL that provide monetary rewards as compensation to incentivize worker participation to provide datasets.** 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 so that the mechanisms can better utilize the available budget). 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 (as we demonstrated in our experiments). **Thus, this paper aims to address these two 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.

Technical Novelty: 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' reputations.

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.3 Reviewer aZoM

Experiments on Complex Task: We further conduct experiments on the dataset CIFAR-10, a image dataset, and the corresponding reputation and the global model accuracy with different number of types (the experiment result w.r.t. the number of requesters will be added into the appendix of the revision) are shown in Table 9 and Table 10. We can observe that our proposed mechanisms always significantly outperform the baselines.

Table 9: Overall Reputation with Different Number of Types

	Dadaset	#Types	BFL(CO/NC)	RRAFL	RanPri
Coop.	CIFAR-10	6	**14.72**	8.05	1.67
		8	**20.70**	11.52	1.90
		10	**23.80**	12.35	1.80
		12	**29.12**	10.42	1.77
Non-coop.	CIFAR-10	6	**17.80**	4.82	2.00
		8	**18.30**	6.02	1.77
		10	**18.57**	9.70	1.76
		12	**17.65**	4.65	1.80

Table 10: Average Global Accuracy with Different Number of Types

	Dadaset	#Types	BFL(CO/NC)	RRAFL	RanPri
Coop.	CIFAR-10	6	**0.484**	0.391	0.222
		8	**0.516**	0.436	0.233
		10	**0.527**	0.456	0.209
		12	**0.540**	0.420	0.217
Non-coop.	CIFAR-10	6	**0.519**	0.292	0.228
		8	**0.487**	0.309	0.218
		10	**0.511**	0.429	0.208
		12	**0.518**	0.393	0.205

Table 11: 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

Heterogeneity: We agree that the incompatibility group modeling is a bit simplified and it can be modeled in a more complex way. We want to note that the consideration of heterogeneous workers in our paper can be used to model a large array of real-world applications. As mentioned in the introduction, in FL where communication channels are utilized to update global model parameters with workers, it is common for workers to experience disconnections due to hardware and bandwidth limitations. This is particularly relevant for workers using mobile devices in congested wireless communication channels [3]. To improve connectivity, FL systems often assign limited channels to workers in order to prevent communication congestion. As a result, a worker allocated to a particular requester must maintain the connection, and it becomes necessary to avoid allocating two workers operating on the same channel to the same requester due to potential interference. In our paper, workers operating on different channels can be categorized into different types. In [3,11], it has been emphasized that neglecting the consideration of channel conflict among workers in FL can lead to communication congestion, which in turn can significantly reduce the accuracy of the global model. Moreover, as no previous literature considers these constraints for budget-constrained incentive mechanism design in FL, we are the first to incorporate such compatibility into the worker selection mechanism. Indeed, this is a highly challenging task even though we represent worker heterogeneity through groups, as it involves designing mechanisms while considering both multiple requesters and budget and compatibility constraints.

For scalability: Table 7 and Table 8 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 11 presents the overall reputation and accuracy results. It is evident that the performance of GreedyPri is significantly inferior to that of our proposed mechanisms.

Cold start: Table 12 presents the average global accuracy on the MNIST dataset (the results for other dataset will be added in the revision) in the cold start scenario, where all workers are assigned the same reputation. It is evident that our mechanisms remain robust in this case. Specifically, the average global accuracy of our mechanisms decreases by 4.51% on average, while for RRAFL, it decreases by 12.79% on average. As PanPri utilizes the random method, its performance is alos extremely poor in this scenario.

B.4 Reviewer C21R

Heterogeneity:

(1) Firstly, we want to note that the consideration of heterogeneous workers in our paper can be used to model a large array of real-world applications. As shown in the introduction, in FL using communication channels to update model parameters, it is necessary to avoid allocating two workers operating on the same channel to the same requester due to potential interference [3,11]. In [3,11], it has been emphasized that neglecting the consideration of channel conflict among workers in FL can lead to communication congestion, which in turn can significantly affect the accuracy of the global model.

(2) We also present the performance of two benchmarks, RRAFL [40] (state-of-the-art) and RanPri, which do not take into account

Table 12: Average Global Accuracy on MNIST under Cold Start (the numbers in parentheses indicate the decrease in accuracy compared to Table 4 and Table 5.)

	#Types	BFL(CO/NC)	RRAFL	RanPri
Coop.	6	**0.834 (0.045)**	0.651 (0.185)	0.494 (-0.073)
	8	**0.816 (0.057)**	0.715 (0.127)	0.439 (0.125)
	10	**0.818 (0.052)**	0.771 (0.060)	0.474 (-0.002)
	12	**0.831 (0.002)**	0.715 (0.113)	0.485 (-0.004)
Non-coop.	6	**0.799 (0.070)**	0.471 (0.221)	0.469 (0.162)
	8	**0.777 (0.090)**	0.645 (0.021)	0.466 (0.066)
	10	**0.821 (0.033)**	0.611 (0.098)	0.489 (0.123)
	12	**0.830 (0.020)**	0.596 (0.160)	0.479 (0.092)
	#Req	BFL(CO/NC)	RRAFL	RanPri
Coop.	4	**0.827 (0.087)**	0.771 (0.070)	0.563 (0.192)
	6	**0.824 (0.050)**	0.696 (0.143)	0.464 (-0.006)
	8	**0.814 (0.036)**	0.718 (0.125)	0.461 (-0.027)
	10	**0.815 (0.027)**	0.637 (0.202)	0.437 (0.005)
Non-coop.	4	**0.809 (0.091)**	0.578 (0.081)	0.484 (0.261)
	6	**0.812 (0.042)**	0.549 (0.139)	0.474 (-0.008)
	8	**0.810 (0.014)**	0.547 (0.134)	0.441 (0.107)
	10	**0.811 (0.006)**	0.501 (0.167)	0.406 (0.015)

the heterogeneity of workers. It is evident that these benchmarks exhibit significantly poorer overall reputation and accuracy compared to our proposed mechanisms.

B.5 Reviewer BWoL

Experiments on Complex Task: We further conduct experiments on the dataset CIFAR-10, a image dataset, and the corresponding reputation and the global model accuracy with different number of types (the experiment result w.r.t. the number of requesters will be added into the appendix of the revision) are shown in Table 9 and Table 10. We can observe that our proposed mechanisms always significantly outperform the baselines.

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