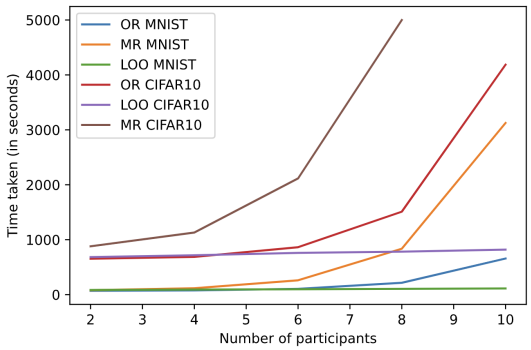

Figure 3: Comparison of Time Complexity.

A Appendix

A.1 Experimental setup

The injected noise is symmetric, meaning corrupted labels are flipped to any other label uniformly at random, and the noise ratios are linearly spaced between 0 and 1; when using 2 clients, their sets have $[0\%, 50\%]$ noise, when using 4 participants the noise rates are $[0\%, 0.25\%, 0.5\%, 0.75\%]$ and so on. Experiments are run with 5 seeds (which remain consistent across the evaluation of different methods), and we only report mean values since variance was negligible. As we are more interested in the cross-silo setting, we assume full client participation in every round.

To compute OR-SV, we use the adjusted version of OR presented in [30], which leads to a faster and more accurate approximation by rearranging terms in the calculation.

For $\lambda - MR$, λ is set to 0.8, per the original paper. Federated Shapley is computed via the same MR approximation, but without the time-decay. To compute round-level LOO, we measure each client's marginal contribution at every round, and sum up the per-round LOO both without weights and multiplied by the value of the current round, i.e., the second-round LOO counts twice as much as the first). The Reputation metric is the average of the Heaviside function applied to the LOO. To arrive at OR-LC, notice that if we approximate all the pseudo-models, we can evaluate them to formulate the LP problem constraints, which is trivial to solve. Since the LC is not unique, we report the first imputation lying in the LC found by the LP solver.

Our setup on MNIST largely follows the setup of [30]. The model used is a two-layer MLP with 64 hidden units and Dropout ($p = 0.5$). No preprocessing is done apart from scaling the images to the $[0, 1]$ range. The number of training rounds is set to 5, and the number of local epochs to 10. Local optimization uses SGD with momentum $m = 0.5$ and $lr = 0.01$.

For the CIFAR-10 experiment, the model used is adapted of the fast ResNet-9 presented in [31]; the architecture is the same but without Batch Norm, and optimization is simplified by training the local models using Adam[32] with learning rate $1e - 3$. The rounds and local epochs remain the same as before. The multi-round approximations of λ-MR and FedShapley need the server to store 1024 ResNet models, causing a crash, even for our lightweight model (but highlighting their memory footprint).

A.2 Time complexity of experiments

We can also examine the computational cost for each family of methods across the two datasets and the number of participants. As expected, MR methods are much more expensive and almost exponential to the number of participants due to high inference cost, but for the more complex CIFAR-10 training, the training itself dominates.

Table 1: Additional results on MNIST; Acc is the global model accuracy. Max Dif is the maximum distance between elements of the payoff vector. Distance refers to the Euclidean distance from the corresponding uniform vector. The time t is given in seconds. B is the total budget allocated by the un-normalized payoffs, and the only metric calculated before they are normalized.

Participants	2					4					6					8					10				
	Acc	Max Dif	Dist	t	B	Acc	Max Dif	Dist	t	B	Acc	Max Dif	Dist	t	B	Acc	Max Dif	Dist	t	B	Acc	Max Dif	Dist	t	B
OR - Shapley	94.9	0.92	0.65	69	0.9	93.31	0.93	0.7	76	0.8	92.29	0.68	0.55	105	0.8	91.6	0.58	0.54	215	0.5	90.93	0.5	0.49	657	0.8
OR - LC		0.74	0.52		1		0.53	0.5		0.9		0.37	0.41		0.9		0.29	0.36		0.9		0.24	0.33		0.9
LOO no weights	94.95	**0.98**	**0.69**	84	2.4	93.3	**0.98**	**0.84**	91	1.3	92.29	**1.04**	**0.9**	97	0.4	91.6	**0.82**	**0.73**	105	0.2	90.9	0.56	0.52	112	0.1
LOO lin. weights		**0.98**	**0.69**		9.6		**0.98**	**0.84**		5.7		0.95	0.84		1.9		0.78	0.7		0.8		**0.7**	**0.63**		0.4
Reputation		0.22	0.15		1.6		0.18	0.18		2.9		0.2	0.2		3.2		0.15	0.16		5.4		0.14	0.18		6
λ-MR	94.85	0.44	0.31	80	3.4	93.34	0.42	0.3	116	3.4	92.35	0.35	0.28	261	3.3	91.62	0.32	0.29	836	3.4	90.95	0.29	0.28	3125	3.3
Federated Shapley		0.56	0.39		4.2		0.52	0.38		4.1		0.41	0.33		4		0.37	0.33		4		0.32	0.31		4

Table 2: Results on CIFAR-10; Acc is the global model accuracy. Max Dif is the maximum distance between elements of the payoff vector. Distance refers to the Euclidean distance from the corresponding uniform vector. The time t is given in seconds. B is the total budget allocated by the un-normalised payoffs, and the only metric calculated before they are normalized.

Participants	2					4					6					8					10				
	Acc	Max Dif	Dist	t	B	Acc	Max Dif	Dist	t	B	Acc	Max Dif	Dist	t	B	Acc	Max Dif	Dist	t	B	Acc	Max Dif	Dist	t	B
OR - Shapley	0.84	0.62	0.44	653	0.71	0.78	0.46	0.34	687	0.68	0.73	0.42	0.34	863	0.63	0.68	0.46	0.42	1510	0.55	0.62	0.46	0.43	4186	0.53
OR - LC		0.31	0.22		0.84		0.41	0.33		0.78		0.3	0.24		0.73		0.23	0.22		0.68		0.17	0.21		0.62
LOO no weights	0.837	**0.94**	**0.66**	683	1.4	0.78	**0.55**	**0.41**	717	1.4	0.74	0.38	0.34	760	1.1	0.68	0.33	0.33	782	1.05	0.65	0.35	0.4	819	0.68
LOO lin. weights		0.74	0.52		6.1		0.47	0.35		5.84		0.31	0.27		4.6		0.29	0.27		4.3		**0.29**	**0.32**		2.9
Reputation		0.25	0.17		1.6		0.11	0.1		3.4		0.125	0.14		4.8		0.17	0.16		5.8		0.14	0.16		7
λ-MR	0.832	0.29	0.2	880	3.3	0.78	0.46	0.35	1130	3.4	0.74	0.48	**0.4**	2116	3.36	0.67	**0.5**	**0.46**	5000	3.36	-	-	-	-	-
Federated Shapley		0.3	0.21		3.2		0.45	0.34		2.9		**0.49**	**0.4**		2.62		**0.5**	**0.46**		2.28		-	-		

A.3 Additional metrics from experiments

Apart from the CE values themselves, we present here the final global accuracy since the test accuracy is our utility function, the maximum difference between two elements in every payoff vector, the Euclidean distance between each payoff vector and an uninformative equal split of the value, and the total computation time. Since we are not concerned with a final model performance, we use the accuracy on the test set as the utility function and do not separate a validation set.

References

[1] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data, 2017.

[2] Andrew Hard, Chloé M Kiddon, Daniel Ramage, Francoise Beaufays, Hubert Eichner, Kanishka Rao, Rajiv Mathews, and Sean Augenstein. Federated learning for mobile keyboard prediction, 2018. URL `https://arxiv.org/abs/1811.03604`.

[3] MELLODDY. MELLODDY: Machine Learning Ledger Orchestration For Drug Discovery, 2019. URL `https://www.melloddy.eu/`.

[4] The Substra Foundation. AI on clinical data : Healthchain consortium, 2019. URL `https://www.substra.ai/en/healthchain-project`.

[5] Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurelien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G. L. D'Oliveira, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrian Gascan, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaid Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konea, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrade Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Azgar, Rasmus Pagh, Mariana Raykova, Hang Qi, Daniel Ramage, Ramesh Raskar, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun, Ananda Theertha Suresh, Florian Tramar, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, and Sen Zhao. Advances and Open Problems in Federated Learning. *arXiv:1912.04977 [cs, stat]*, December 2019. URL `http://arxiv.org/abs/1912.04977`. arXiv: 1912.04977.

[6] Lloyd S Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2(28): 307–317, 1953.

[7] T. Ichiishi and K. Shell. *Game Theory for Economic Analysis*. Economic theory, econometrics, and mathematical economics. Academic Press, 1983. ISBN 9780123701800.

[8] B. Peleg and P. Sudholter. *Introduction to the theory of cooperative games*. Theory and Decision Library C. Springer-Verlag Berlin Heidelberg, 2007. ISBN 9783540729440.

[9] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1175–1191, 2017.

[10] Amirata Ghorbani and James Zou. Data Shapley: Equitable Valuation of Data for Machine Learning. *arXiv:1904.02868 [cs, stat]*, June 2019. arXiv: 1904.02868.

[11] Amirata Ghorbani, Michael P. Kim, and James Zou. A Distributional Framework for Data Valuation. *arXiv:2002.12334 [cs, stat]*, February 2020. URL `http://arxiv.org/abs/2002.12334`. arXiv: 2002.12334.

[12] Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gurel, Bo Li, Ce Zhang, Dawn Song, and Costas Spanos. Towards Efficient Data Valuation Based on the Shapley Value. *arXiv:1902.10275 [cs, stat]*, August 2019. arXiv: 1902.10275.

[13] Tianhao Wang, Johannes Rausch, Ce Zhang, Ruoxi Jia, and Dawn Song. A Principled Approach to Data Valuation for Federated Learning. *arXiv:2009.06192 [cs, stat]*, September 2020. URL `http://arxiv.org/abs/2009.06192`. arXiv: 2009.06192.

[14] Shuyue Wei, Yongxin Tong, Zimu Zhou, and Tianshu Song. *Efficient and Fair Data Valuation for Horizontal Federated Learning*. 2020.

[15] Tianshu Song, Yongxin Tong, and Shuyue Wei. Profit allocation for federated learning. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2577–2586, 2019. doi: 10.1109/BigData47090.2019.9006327.

[16] Tom Yan and Ariel D Procaccia. If You Like Shapley Then You'll Love the Core. In *AAAI 2021*, 2021.

[17] Harry Cai, Daniel Rueckert, and Jonathan Passerat-Palmbach. 2cp: Decentralized protocols to transparently evaluate contributivity in blockchain federated learning environments, 2020.

[18] Shuaicheng Ma, Yang Cao, and Li Xiong. Transparent Contribution Evaluation for Secure Federated Learning on Blockchain. *arXiv:2101.10572 [cs]*, January 2021. arXiv: 2101.10572.

[19] Yuan Liu, Shuai Sun, Zhengpeng Ai, Shuangfeng Zhang, Zelei Liu, and Han Yu. Fedcoin: A peer-to-peer payment system for federated learning, 2020.

[20] Jierui Lin, Min Du, and Jian Liu. Free-riders in federated learning: Attacks and defenses. *arXiv preprint arXiv:1911.12560*, 2019.

[21] Yann Fraboni, Richard Vidal, and Marco Lorenzi. Free-rider attacks on model aggregation in federated learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1846–1854. PMLR, 2021.

[22] Rachael Hwee Ling Sim, Yehong Zhang, Mun Choon Chan, and Bryan Kian Hsiang Low. Collaborative Machine Learning with Incentive-Aware Model Rewards. In *International Conference on Machine Learning*, pages 8927–8936. PMLR, November 2020.

[23] Lingjuan Lyu, Xinyi Xu, Qian Wang, and Han Yu. *Collaborative Fairness in Federated Learning*. 2020.

[24] Lingjuan Lyu, Yitong Li, Karthik Nandakumar, Jiangshan Yu, and Xingjun Ma. How to Democratise and Protect AI: Fair and Differentially Private Decentralised Deep Learning. *IEEE Trans. Dependable and Secure Comput.*, pages 1–1, 2020. ISSN 1545-5971, 1941-0018, 2160-9209. arXiv: 2007.09370.

[25] Jingfeng Zhang, Cheng Li, Antonio Robles-Kelly, and Mohan Kankanhalli. Hierarchically Fair Federated Learning. *arXiv:2004.10386 [cs, stat]*, May 2020. arXiv: 2004.10386.

[26] About substra. URL `https://www.substra.ai/en/about-substra-foundation`.

[27] J. Kang, Z. Xiong, D. Niyato, S. Xie, and J. Zhang. Incentive Mechanism for Reliable Federated Learning: A Joint Optimization Approach to Combining Reputation and Contract Theory. *IEEE Internet of Things Journal*, 6(6):10700–10714, December 2019. ISSN 2327-4662. doi: 10.1109/JIOT.2019.2940820. Conference Name: IEEE Internet of Things Journal.

[28] Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*.

[29] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 (canadian institute for advanced research). 2012. URL `http://www.cs.toronto.edu/~kriz/cifar.html`.

[30] Le Quang Anh. Time-efficient algorithms to approximate the shapley values for horizontal enterprise federated learning. URL `https://github.com/Toefinder/Shapley-Algorithms-Federated-Learning-`.

[31] How to train your resnet. URL `https://myrtle.ai/learn/how-to-train-your-resnet-8-bag-of-tricks/`.

[32] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.