Proceedings of Machine Learning Research vol 178:1–5, 2022 35th Annual Conference on Learning Theory

Efficient decentralized multi-agent learning in asymmetric queuing systems (extended abstract)

Daniel Freund DFREUND@MIT.EDU
Thodoris Lykouris LYKOURIS@MIT.EDU
Wentao Weng WWENG@MIT.EDU
Massachusetts Institute of Technology

Editors: Po-Ling Loh and Maxim Raginsky

Motivated by packet routing in computer networks and resource allocation in cognitive radio, bipartite queuing systems have risen as a canonical setting to capture carryover effects in sequential learning. In this setting, there are N agents and K servers. Each agent i receives jobs with a fixed arrival rate λ_i and selects a server j to route their job. The server selects (at most) one of the requesting agents i and successfully serves her job with service rate $\mu_{i,j}$. Any non-served job returns to its respective agent and is stored in a queue in front of her.

Although this kind of queuing system has long been a standard approach to model service systems, a learning lens has only recently been introduced to this context. In particular, Krishnasamy et al. (2016, 2021) introduced this line of work by studying a centralized view of the problem where a learner is allowed to jointly control all agents (there, agents correspond to different classes of jobs). That said, many queuing systems exhibit a decentralized nature, in which agents do not have the ability to communicate. Very recently, two works initiated the study of decentralized multi-agent learning in queuing systems for the symmetric case where the service rates are only affected by the server j, i.e., $\mu_{i,j} = \mu_j$ for all agents i. Gaitonde and Tardos (2020) studied the quality of outcomes when agents are strategic and use no-regret learning algorithms to maximize their individual welfare and showed that the system only stabilizes if it has twice as much capacity as a central controller requires. Closer to our work, Sentenac et al. (2021) provided a collaborative scheme that, when followed by all agents, provides bounded average-time queue lengths, i.e., it stabilizes the system for any positive traffic slackness without online communication or knowledge of the service rates. Despite providing the first decentralized learning algorithm for bipartite queuing systems, the algorithm of Sentenac et al. (2021) does not scale well to systems with even a moderate number of servers K, requires significant initial coordination to hardwire subsequent communication, and needs the system to be symmetric.

In this work, we design a simple learning the first decentralized learning algorithm for online queuing systems that achieves the following desiderata. It is *sample-efficient* in the sense that its queue-length guarantee is polynomial in the system parameters K and N. It is *computationally efficient* as its running time is linear in K and independent of N. It is *fully decentralized* in the sense that all agents use exactly the same simple algorithm without needing to have a unique identifier or shared randomness. Finally, our results hold for any *asymmetric* bipartite queuing system. In fact, our algorithms even extend to a dynamic setting where queues can join the system or depart: we prove that a dynamic version of our algorithm adapts to a changing set of queues as long as there is a known upper bound on their number at any time. Along the way, we provide the first UCB-based algorithm even for the centralized case of the problem, which replaces the need for *forced exploration* and resolves an open question by Krishnasamy et al. (2016, 2021). [1]

1. Extended abstract. Full version appears as [Freund et al. (2022), v1]

References

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

Orly Avner and Shie Mannor. Concurrent bandits and cognitive radio networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 66–81. Springer, 2014.

Mohsen Bayati, Balaji Prabhakar, Devavrat Shah, and Mayank Sharma. Iterative scheduling algorithms. In *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*, pages 445–453. IEEE, 2007.

Dimitri P Bertsekas. The auction algorithm: A distributed relaxation method for the assignment problem. *Annals of operations research*, 14(1):105–123, 1988.

Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

Ilai Bistritz and Amir Leshem. Game of thrones: Fully distributed learning for multiplayer bandits. *Mathematics of Operations Research*, 46(1):159–178, 2021.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Etienne Boursier and Vianney Perchet. Sic-mmab: synchronisation involves communication in multiplayer multi-armed bandits. *Advances in Neural Information Processing Systems*, 32, 2019.

Lawrence Brown, Noah Gans, Avishai Mandelbaum, Anat Sakov, Haipeng Shen, Sergey Zeltyn, and Linda Zhao. Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American statistical association*, 100(469):36–50, 2005.

Sébastien Bubeck, Thomas Budzinski, and Mark Sellke. Cooperative and stochastic multi-player multi-armed bandit: Optimal regret with neither communication nor collisions. In *Conference on Learning Theory*, pages 821–822. PMLR, 2021.

Jinsheng Chen, Jing Dong, and Pengyi Shi. A survey on skill-based routing with applications to service operations management. *Queueing Systems*, 96(1):53–82, 2020.

Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit: General framework and applications. In *International conference on machine learning*, pages 151–159. PMLR, 2013.

Tuhinangshu Choudhury, Gauri Joshi, Weina Wang, and Sanjay Shakkottai. Job dispatching policies for queueing systems with unknown service rates. In *Proceedings of the Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, pages 181–190, 2021.

Gabrielle Demange, David Gale, and Marilda Sotomayor. Multi-item auctions. *Journal of political economy*, 94(4):863–872, 1986.

Daniel Freund, Thodoris Lykouris, and Wentao Weng. Efficient decentralized multi-agent learning in asymmetric queuing systems, 2022.

Yi Gai, Bhaskar Krishnamachari, and Rahul Jain. Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation. In *2010 IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*, pages 1–9. IEEE, 2010.

Jason Gaitonde and Eva Tardos. Virtues of patience in strategic queuing systems. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, 2021.

Jason Gaitonde and Éva Tardos. Stability and learning in strategic queuing systems. In *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020. ISBN 978-1-4503-7975-5. doi: 10.1145/3391403.3399491. URL https://dl.acm.org/doi/10.1145/3391403.3399491.

Abhinav Gupta, Xiaojun Lin, and Rayadurgam Srikant. Low-complexity distributed scheduling algorithms for wireless networks. *IEEE/ACM Transactions on Networking*, 17(6):1846–1859, 2009.

Itay Gurvich and Ward Whitt. Scheduling flexible servers with convex delay costs in many-server service systems. *Manufacturing & Service Operations Management*, 11(2):237–253, 2009.

Meena Jagadeesan, Alexander Wei, Yixin Wang, Michael Jordan, and Jacob Steinhardt. Learning equilibria in matching markets from bandit feedback. *Advances in Neural Information Processing Systems*, 34, 2021.

Libin Jiang and Jean C. Walrand. A distributed CSMA algorithm for throughput and utility maximization in wireless networks. *IEEE/ACM Trans. Netw.*, 18(3):960–972, 2010. doi: 10.1109/TNET.2009.2035046. URL https://doi.org/10.1109/TNET.2009.2035046.

Chi Jin, Qinghua Liu, Yuanhao Wang, and Tiancheng Yu. V-learning–a simple, efficient, decentralized algorithm for multiagent rl. *arXiv preprint arXiv:2110.14555*, 2021.

Dileep Kalathil, Naumaan Nayyar, and Rahul Jain. Decentralized learning for multiplayer multi-armed bandits. *IEEE Transactions on Information Theory*, 60(4):2331–2345, 2014.

Hsu Kao, Chen-Yu Wei, and Vijay Subramanian. Decentralized cooperative reinforcement learning with hierarchical information structure. *arXiv preprint arXiv:2111.00781*, 2021.

Subhashini Krishnasamy, Rajat Sen, Ramesh Johari, and Sanjay Shakkottai. Regret of queueing bandits. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 1669–1677, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/430c3626b879b4005d41b8a46172e0c0-Abstract.html.

Subhashini Krishnasamy, PT Akhil, Ari Arapostathis, Rajesh Sundaresan, and Sanjay Shakkottai. Augmenting max-weight with explicit learning for wireless scheduling with switching costs. *IEEE/ACM Transactions on Networking*, 26(6):2501–2514, 2018.

Subhashini Krishnasamy, Rajat Sen, Ramesh Johari, and Sanjay Shakkottai. Learning unknown service rates in queues: A multiarmed bandit approach. *Oper. Res.*, 69(1):315–330, 2021. doi: 10.1287/opre.2020.1995. URL https://doi.org/10.1287/opre.2020.1995.

Qingkai Liang and Eytan Modiano. Minimizing queue length regret under adversarial network models. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2(1): 1–32, 2018.

Lydia T Liu, Feng Ruan, Horia Mania, and Michael I Jordan. Bandit learning in decentralized matching markets. *Journal of Machine Learning Research*, 22(211):1–34, 2021.

Gábor Lugosi and Abbas Mehrabian. Multiplayer bandits without observing collision information. *Mathematics of Operations Research*, 2021.

Thodoris Lykouris, Vasilis Syrgkanis, and Éva Tardos. Learning and efficiency in games with dynamic population. In *Proceedings of the Twenty-Seventh Annual Symposium on Discrete Algorithms (SODA)*, pages 120–129. SIAM, 2016. doi: 10.1137/1.9781611974331.ch9. URL https://doi.org/10.1137/1.9781611974331.ch9.

Siva Theja Maguluri and R Srikant. Heavy traffic queue length behavior in a switch under the maxweight algorithm. *Stochastic Systems*, 6(1):211–250, 2016.

Avishai Mandelbaum and Alexander L. Stolyar. Scheduling flexible servers with convex delay costs: Heavy-traffic optimality of the generalized cμ-rule. *Operations Research*, 52(6):836–855, Dec 2004. ISSN 0030-364X, 1526-5463. doi: 10.1287/opre.1040.0152.

Abbas Mehrabian, Etienne Boursier, Emilie Kaufmann, and Vianney Perchet. A practical algorithm for multiplayer bandits when arm means vary among players. In *International Conference on Artificial Intelligence and Statistics*, pages 1211–1221. PMLR, 2020.

Michael J Neely, Scott T Rager, and Thomas F La Porta. Max weight learning algorithms for scheduling in unknown environments. *IEEE Transactions on Automatic Control*, 57(5):1179–1191, 2012.

Jonathan Rosenski, Ohad Shamir, and Liran Szlak. Multi-player bandits–a musical chairs approach. In *International Conference on Machine Learning*, pages 155–163. PMLR, 2016.

Flore Sentenac, Etienne Boursier, and Vianney Perchet. Decentralized learning in online queuing systems. *arXiv preprint arXiv:2106.04228*, 2021.

Devavrat Shah and Milind Kopikare. Delay bounds for approximate maximum weight matching algorithms for input queued switches. In *Proceedings.Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, page 1024–1031. IEEE, 2002. ISBN 978-0-7803-7476-8. doi: 10.1109/INFCOM.2002.1019350. URL http://ieeexplore.ieee.org/document/1019350/.

Devavrat Shah and Jinwoo Shin. Randomized scheduling algorithm for queueing networks. *The Annals of Applied Probability*, 22(1):128–171, 2012.

Devavrat Shah and Damon Wischik. Switched networks with maximum weight policies: Fluid approximation and multiplicative state space collapse. *The Annals of Applied Probability*, 22(1): 70–127, 2012.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*, 2019.

R. Srikant and Lei Ying. *Communication networks: an optimization, control, and stochastic networks perspective*. Cambridge University Press, 2014. ISBN 978-1-107-03605-5.

Thomas Stahlbuhk, Brooke Shrader, and Eytan Modiano. Learning algorithms for minimizing queue length regret. *IEEE Transactions on Information Theory*, 67(3):1759–1781, 2021.

Alexander L. Stolyar. Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic. *The Annals of Applied Probability*, 14(1), Feb 2004. ISSN 1050-5164. doi: 10.1214/aoap/1075828046.

Leandros Tassiulas. Linear complexity algorithms for maximum throughput in radio networks and input queued switches. In *Proceedings. IEEE INFOCOM'98, the Conference on Computer Communications. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Gateway to the 21st Century (Cat. No. 98*, volume 2, pages 533–539. IEEE, 1998.

Leandros Tassiulas and Anthony Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 37(12):1936–1948, Dec 1992. ISSN 00189286. doi: 10.1109/9.182479.

Neil Walton and Kuang Xu. Learning and information in stochastic networks and queues. In *Tutorials in Operations Research: Emerging Optimization Methods and Modeling Techniques with Applications*, pages 161–198. INFORMS, 2021.

Neil S Walton. Two queues with non-stochastic arrivals. *Operations Research Letters*, 42(1):53–57, 2014.

Wentao Weng, Xingyu Zhou, and R. Srikant. Optimal load balancing with locality constraints. *Proc. ACM Meas. Anal. Comput. Syst.*, 4(3):45:1–45:37, 2020. doi: 10.1145/3428330. URL https://doi.org/10.1145/3428330.

Michael M Zavlanos, Leonid Spesivtsev, and George J Pappas. A distributed auction algorithm for the assignment problem. In *2008 47th IEEE Conference on Decision and Control*, pages 1212–1217. IEEE, 2008.