Proceedings of Machine Learning Research vol 134:1–2, 2021

# Cooperative and Stochastic Multi-Player Multi-Armed Bandit: Optimal Regret With Neither Communication Nor Collisions

**Sébastien Bubeck** SEBUBECK@MICROSOFT.COM
**Microsoft Research**

**Thomas Budzinski** THOMAS.BUDZINSKI@ENS-LYON.FR
**UBC**

**Mark Sellke** MSELLKE@STANFORD.EDU
**Stanford University**

## Abstract

We consider the cooperative multi-player version of the stochastic multi-armed bandit problem. The mutiplayer bandit problem with limited communication was first introduced roughly at the same time in Lai et al. (2008); Liu and Zhao (2010); Anandkumar et al. (2011), and has been extensively studied since then Avner and Mannor (2014); Rosenski et al. (2016); Bonnefoi et al. (2017); Lugosi and Mehrabian (2018); Boursier and Perchet (2019); Alatur et al. (2019); Bubeck et al. (2020), with various assumptions on the communication/collisions.

We study the regime where the players cannot communicate but have access to shared randomness. Previously in Bubeck and Budzinski (2020) a strategy for this regime was constructed for two players and three arms, with regret $\tilde{O}(\sqrt{T})$, and with no collisions at all between the players (with very high probability). In this paper we show that these properties (near-optimal regret and no collisions at all) are achievable for any number of players and arms. The previous strategy heavily relied on a 2-dimensional geometric intuition that was difficult to generalize in higher dimensions. We replace it by a tree-based space partition that applies in full generality. At a high level, our partitioning scheme ensures that players do not collide whenever their parameter estimates are close to each other, while allowing them to play optimally except on thin, random slices of the parameter space.

**Keywords:** Multi-armed bandit, multi-agent learning

## Acknowledgments

The third author acknowledges funding from NSF and Stanford graduate fellowships.

. Extended abstract. Full version appears as arxiv:2011.03896, v1

## References

P. Alatur, K. Y. Levy, and A. Krause. Multi-player bandits: The adversarial case. *Journal of Machine Learning Research (JMLR)*, 2019.

A. Anandkumar, N. Michael, A. K. Tang, and A. Swami. Distributed algorithms for learning and cognitive medium access with logarithmic regret. *IEEE Journal on Selected Areas in Communications*, 29(4):731–745, 2011.

O. Avner and S. Mannor. Concurrent bandits and cognitive radio networks. In *ECML/PKDD*, 2014.

R. Bonnefoi, L. Besson, C. Moy, E. Kaufmann, and J. Palicot. Multi-armed bandit learning in iot networks: Learning helps even in non-stationary settings. In *International Conference on Cognitive Radio Oriented Wireless Networks*, pages 173–185. Springer, 2017.

Etienne Boursier and Vianney Perchet. Sic-mmab: synchronisation involves communication in multiplayer multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 12071–12080, 2019.

S. Bubeck, Y. Li, Y. Peres, and M. Sellke. Non-stochastic multi-player multi-armed bandits: Optimal rate with collision information, sublinear without. In *COLT*, 2020.

Sébastien Bubeck and Thomas Budzinski. Coordination without communication: optimal regret in two players multi-armed bandits. In *COLT*, 2020.

L. Lai, H. Jiang, and H. V. Poor. Medium access in cognitive radio networks: A competitive multi-armed bandit framework. In *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 98–102, 2008.

K. Liu and Q. Zhao. Distributed learning in multi-armed bandit with multiple players. *IEEE Transactions on Signal Processing*, 58(11):5667–5681, 2010.

G. Lugosi and A. Mehrabian. Multiplayer bandits without observing collision information. *arXiv preprint arXiv:1808.08416*, 2018.

J. Rosenski, O. Shamir, and L. Szlak. Multi-player bandits - a musical chairs approach. In *ICML*, 2016.