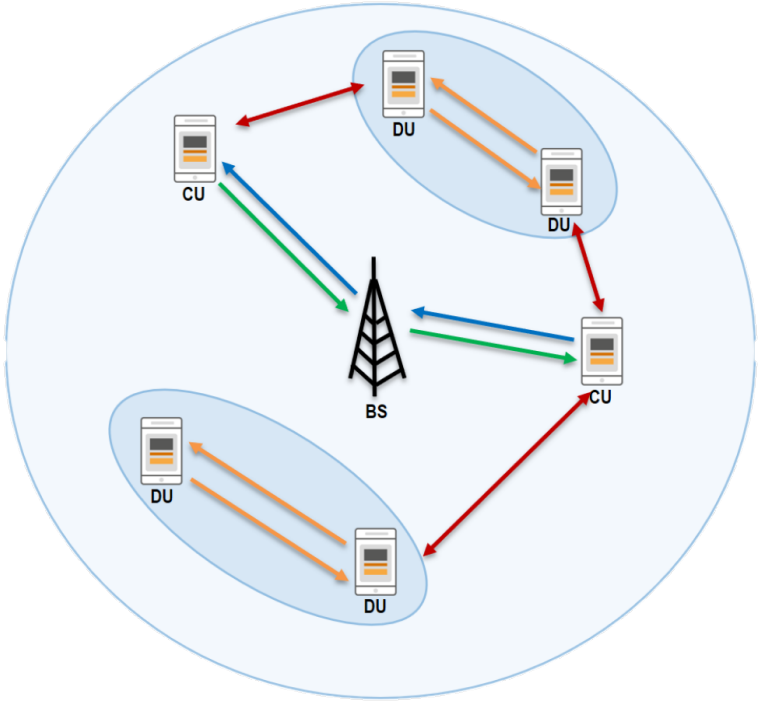
salzburgresearch



Multi-Agent Reinforcement Learning for Resource Allocation in Wireless Network Communications

Sabrina Pochaba, Peter Dorfinger,
Roland Kwitt, Simon Hirlaender

31/03/2026

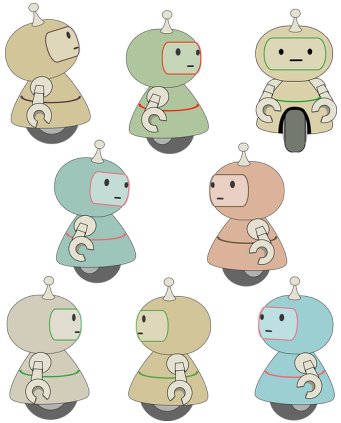
RL4AA'26
Contributed talks

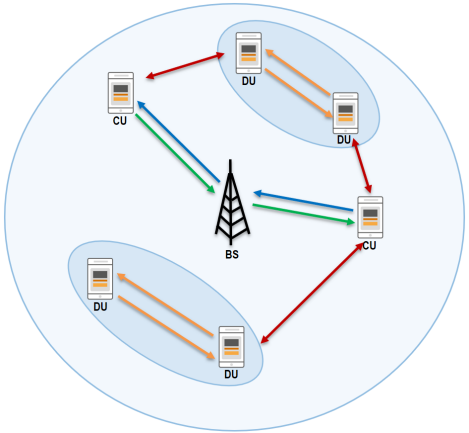
Agenda

- Multi-Agent Reinforcement Learning (MARL)
- MARL Algorithms
- Comparison of Algorithms
- Wireless Network Communication
- Results

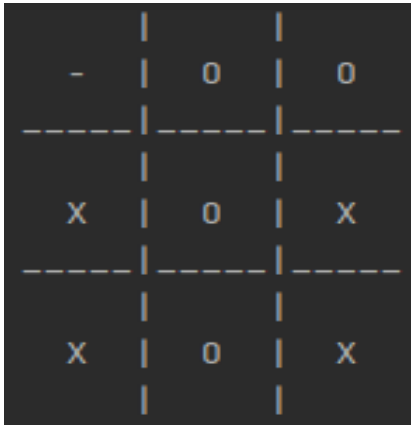
Multi-Agent Reinforcement Learning

Cooperative Games:





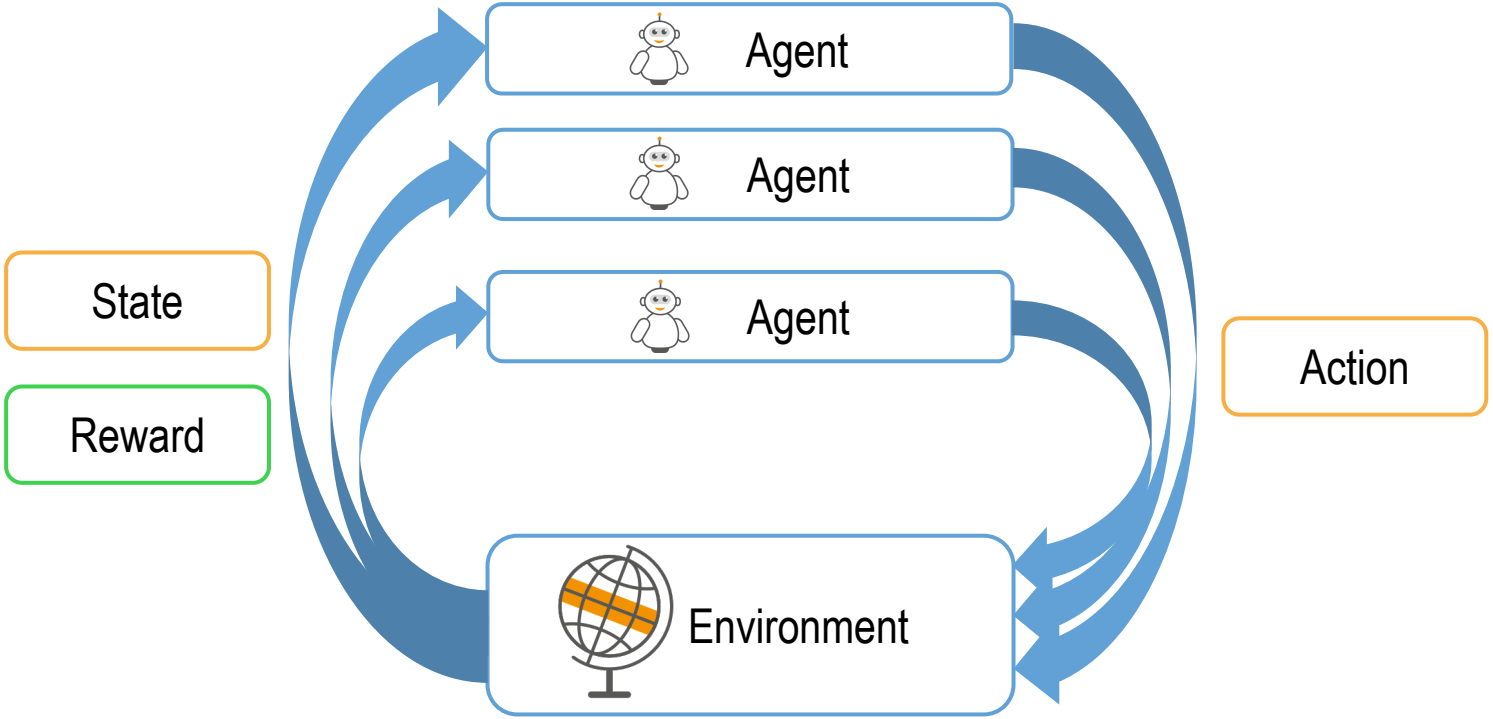
Adversarial Games:



Mixed Games:

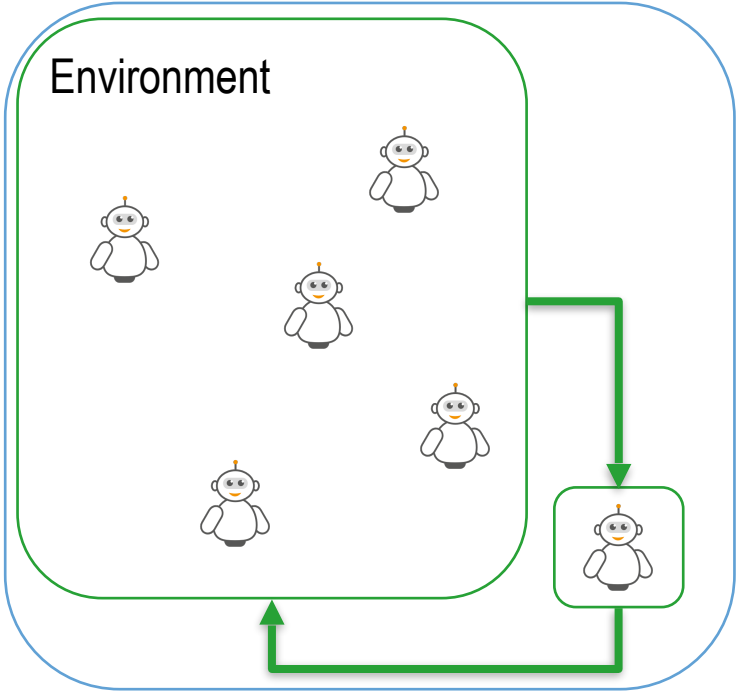
https://www.youtube.com/watch?v=kopoLzvh5jY;
Baker B., Kanitscheier I., Markov T., Wu Y., Emergent Tool Use From Multi-Agent Autocurricula, 2020

Multi-Agent Reinforcement Learning



Challenges of MARL

- Non-unique Learning Goals
- Non-stationarity
- Scalability issues
- Cooperative vs competitive



MARL Algorithms: IQL

- Core idea: each agent learns its own Q-function independently
- Information: local state/ observation, own action, local/ shared reward
- Strength: simple, scalable, easy baseline
- Weakness: other agents are seen as part of the env, causing non-stationarity

- Q-function: $Q_i(s_i, a_i)$

MARL Algorithms: NashQ

- Core idea: learn joint Q-values & choose actions based on Nash equilibrium
- Information: full joint action payoff structure
- Strength: principled treatment of strategic interaction
- Weakness: NE computation is expensive, can be unstable, multiple NEs possible

- Q-function: $Q_i(s_i, a_{NE})$

MARL Algorithms: JAL

- Core idea: each agent learns Q-values over joint actions
- Information: own state + actions of all agents or estimation of other actions
- Strength: explicitly captures interaction between agents
- Weakness: joint Action Space grows combinatorially

- Q-function: $Q_i(s_i, a_1, \ldots, a_n)$

MARL Algorithms: VDN

- Core idea: team value is decomposed into the sum of individual agent Q-valued
- Information: individual utilities combined into one team utility
- Strength: supports cooperative learning while keeping decentralized execution
- Weakness: additive decomposition can be too restrictive

- Q-function: $Q_{total} = \sum_{i} Q_i(s_i, a_i)$

Joint team value is the sum of individual values

Comparison of MARL Algorithms

Criterion	IQL	NashQ	JAL	VDN
Suitable for cooperative tasks				
Sensitivity to non-stationarity				
Coordination capability				
Scalability				
Computational complexity				
Sum up				

Comparison of MARL Algorithms

Criterion	IQL	NashQ	JAL	VDN
Suitable for cooperative tasks				
Sensitivity to non-stationarity				
Coordination capability				
Scalability				
Computational complexity				
Sum up				

Comparison of MARL Algorithms

Criterion	IQL	NashQ	JAL	VDN
Suitable for cooperative tasks				
Sensitivity to non-stationarity				
Coordination capability				
Scalability				
Computational complexity				
Sum up				

Comparison of MARL Algorithms

Criterion	IQL	NashQ	JAL	VDN
Suitable for cooperative tasks				
Sensitivity to non-stationarity				
Coordination capability				
Scalability				
Computational complexity				
Sum up				

Comparison of MARL Algorithms

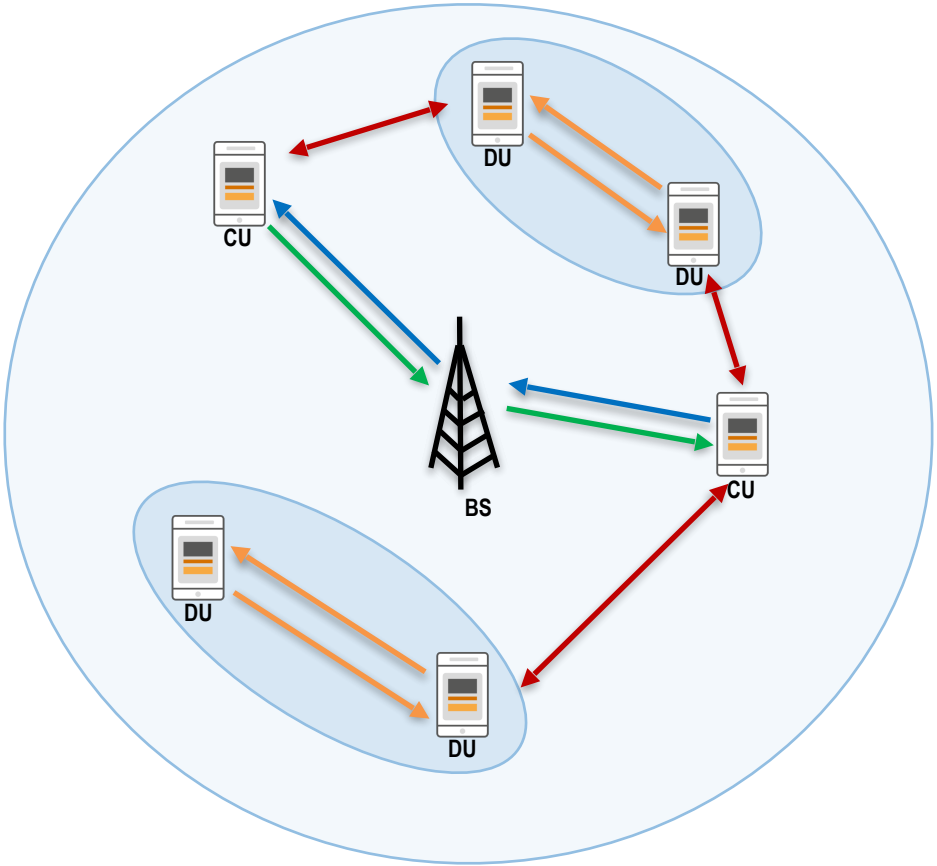
Criterion	IQL	NashQ	JAL	VDN
Suitable for cooperative tasks				
Sensitivity to non-stationarity				
Coordination capability				
Scalability				
Computational complexity				
Sum up				

Comparison of MARL Algorithms

Criterion	IQL	NashQ	JAL	VDN
Suitable for cooperative tasks				
Sensitivity to non-stationarity				
Coordination capability				
Scalability				
Computational complexity				
Sum up	easy, but difficult training	highly depending on NE	coordination vs scalability	strong cooperation, weakly competitive

Cellular and D2D communication

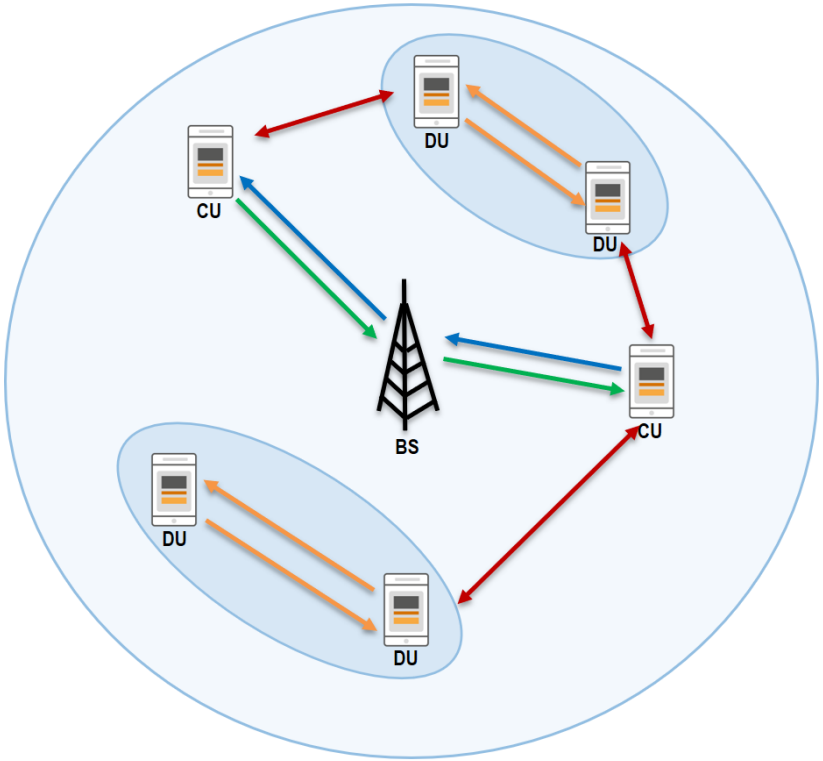
BS = Base Station
CU = Cellular User
DU = D2D User

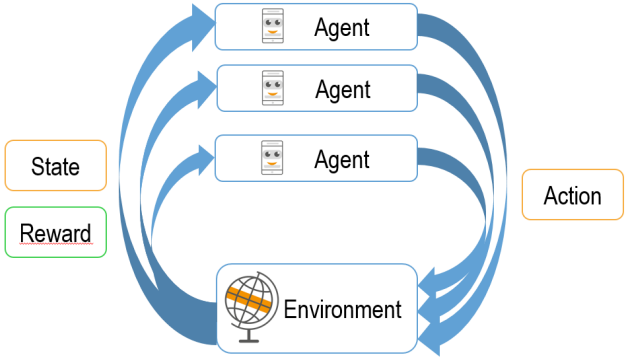


Problem:
Reliable Communication without regulation of BS

Solution:
Multi-Agent Reinforcement Learning

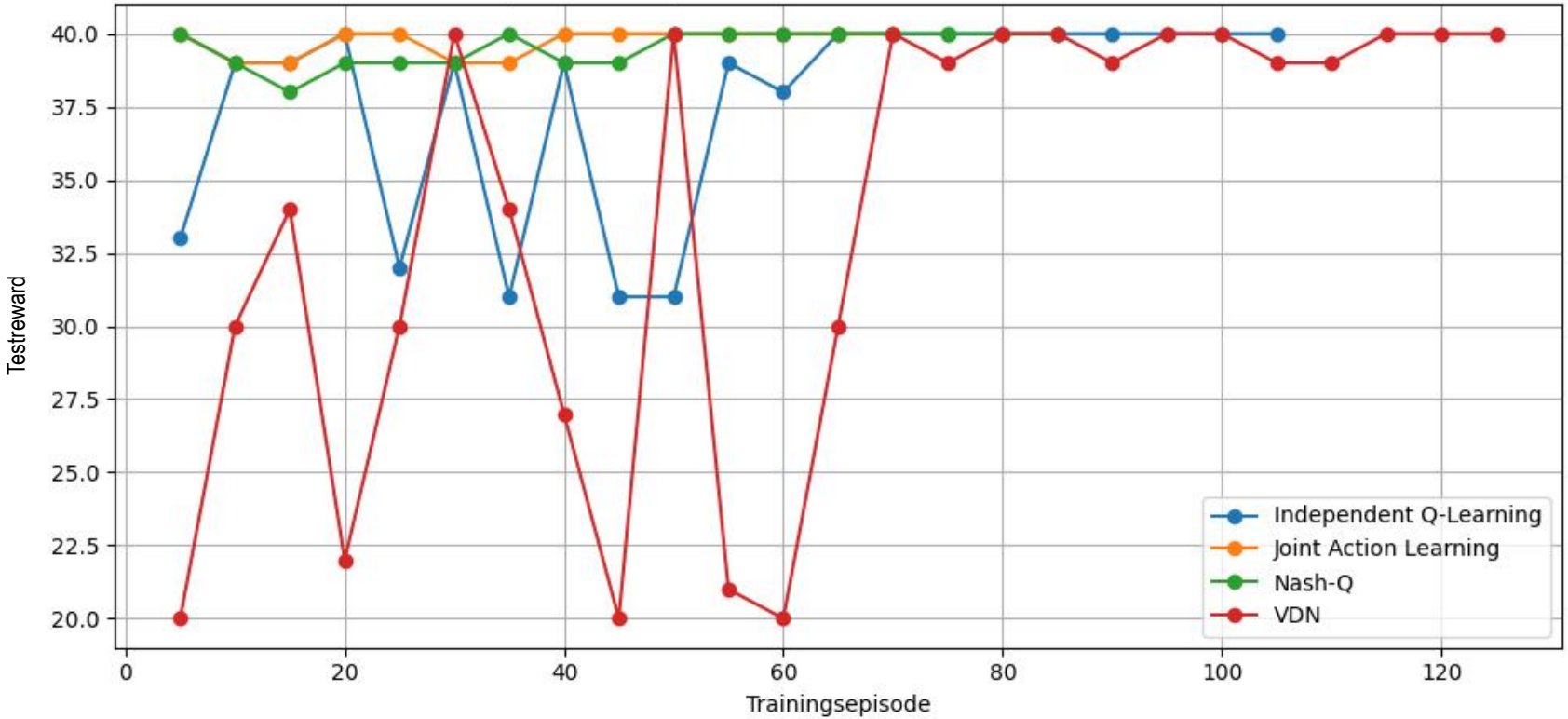
MARL in D2D communication



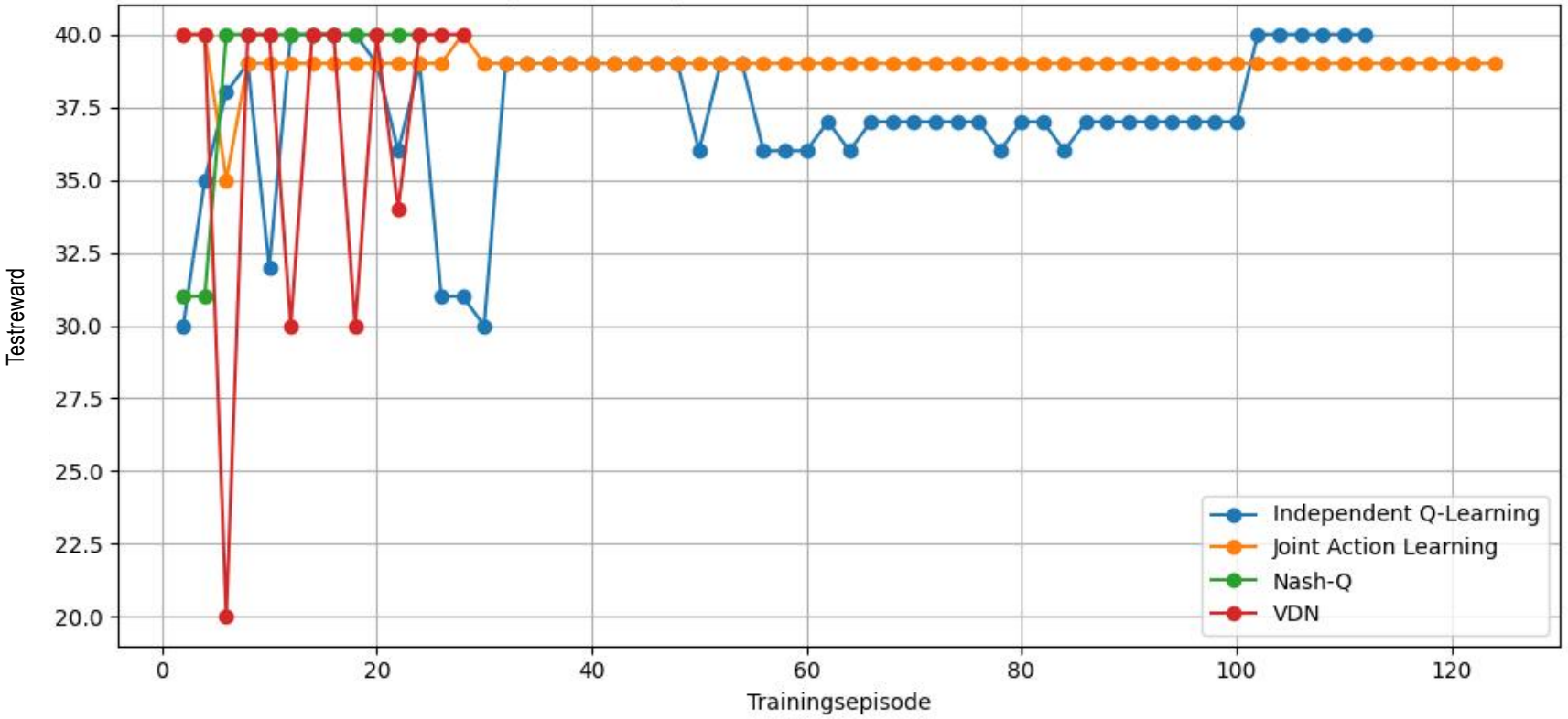


- Action: Choose Channel
- State:
  - Own Channel selection
  - Satisfaction (QoS)
  - Neighbors
  - Channel selection of neighbors
- Reward: Satisfaction (QoS) of all devices

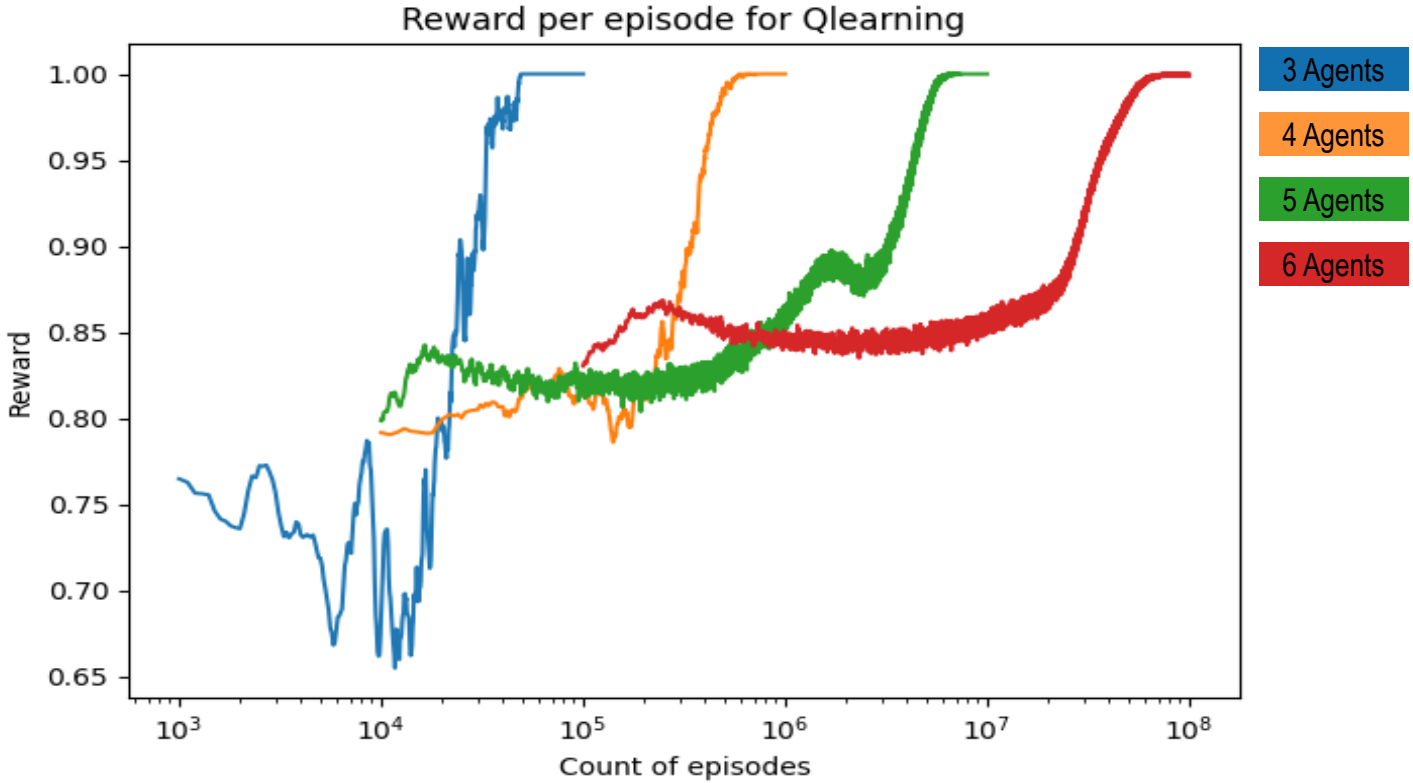
Results (2 Agents, 2 Channels)



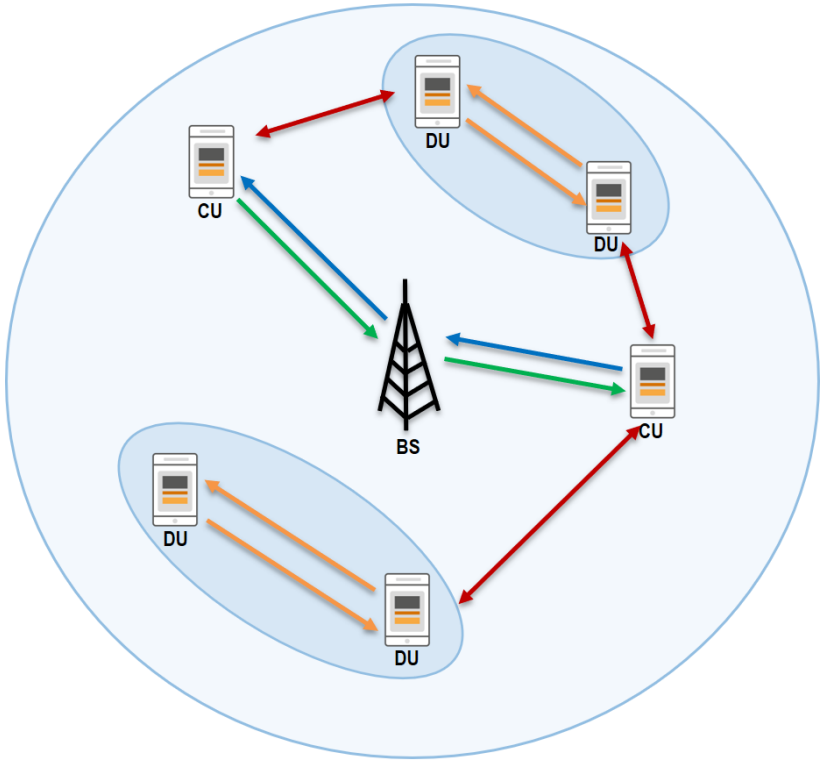
Results (2 Agents, 3 Channels)



Results


Reward per episode for Qlearning
Reward
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
10^3
10^4
10^5
10^6
10^7
10^8
Count of episodes
3 Agents
4 Agents
5 Agents
6 Agents

Conclusion



- MARL Algorithms:
  - IQL: learns quickly, but struggle to stabilize
  - NashQ: powerful, but computationally complex
  - JAL: improves interaction, but high scalability
  - VDN: very strong in cooperative settings, but only there
- Number of Agents increases training time
- Training Many-Agent-Scenarios with different amount of channels
- Assumption: VDN-algorithm outperforms other
- Reduce information
- Movement of devices

Sabrina Pochaba

Salzburg Research Forschungsgesellschaft mbH
Jakob-Haringer-Str. 5/3 | Salzburg, Austria

+43 662 2288-459

sabrina.pochaba@salzburgresearch.at

Questions?

References

- Stefano V. Albrecht, Filippos Christianos, Lukas Schäfer. Multi-Agent Reinforcement Learning: Foundations and Modern Approaches.The MIT Press, 2024
- Junling Hu, Michael P.Wellman. Nash Q-Learning for General-Sum Stochastic games. Journal of Mahcine Learning Research 4 (2003)
- Arash Asadi, Qing Wang, and Vincenzo Mancuso. A survey on device-to-device communication in cellular networks. IEEE Communications Surveys Tutorials,16(4):1801–1
- Khaled B. Letaief, Wei Chen, Yuanming Shi, Jun Zhang, and Ying-Jun Angela Zhang. The roadmap to 6g: Ai em-powered wireless networks. IEEE Communications Magazine, 57(8):84–90, 2019.
- Andreas F. Molisch. Wireless Communications. Wiley Publishing, 2nd edition, 2011.
- Ann Nowé, Peter Vrancx, and De Hauwere Yann-Micha¨el. Game Theory and Multi-agent Reinforcement Learning, pages 441–470. Springer-Verlag Berlin Heidelberg, 2012.
- Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. The MIT Press, second edition, 2018
- Huaqing Zhang and Shanghang Zhang. Multi-Agent Reinforcement Learning, pages 335–346. Springer Singapore, Singapore, 2020.
- Kaiqing Zhang, Zhuoran Yang, and Tamer Basar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. CoRR, abs/1911.10635, 2019.
- Yuan Zhi, Jie Tian, Xiaofang Deng, Jingping Qiao, and Dianjie Lu. Deep reinforcement learning-based resource allocation for d2d communications in heterogeneous cellular networks. Digital Communications and Networks, 2021.