Distributed Resource Allocation for 5G-V2X Communication with Multi-agent Reinforcement Learning

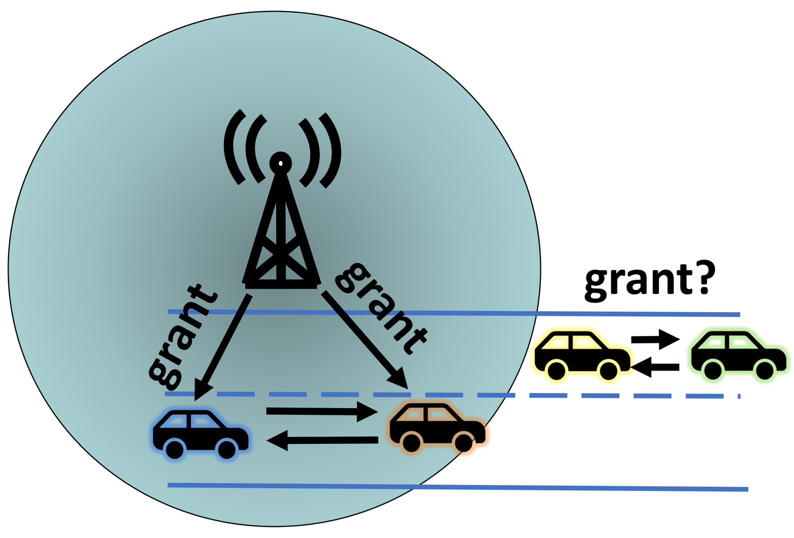
Alperen Gündogan, Volker Pauli, H. Murat Gürsu, Wolfgang Kellerer

alperen.guendogan@tum.de, pauli@nomor.de, murat.guersu@tum.de, wolfgang.kellerer@tum.de

Chair of Communication Networks, Technische Universität München, Germany

Motivation

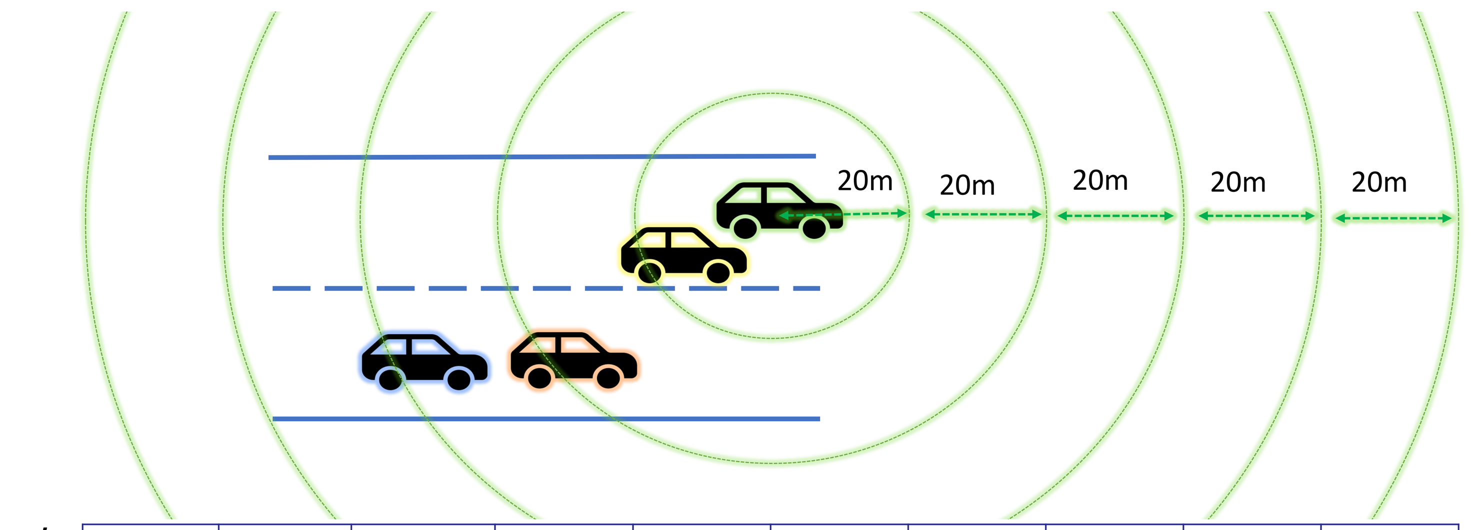
- Millions of death from car accidents[1].
- **Vehicle-to-everything(V2X) Communication**
  - Presence of base station can not be guaranteed.
- **Distributed radio access technologies**
  - Cellular-V2X(C-V2X)
    - **Spatial-Local observation for decisions**
    - Based on Carrier Sensing



- **Problem: Near** vehicles are likely to select the **same** resources[2].
- *Far vehicles use the same resources with multi-agent reinforcement learning(MARL) in congestion scenario.*

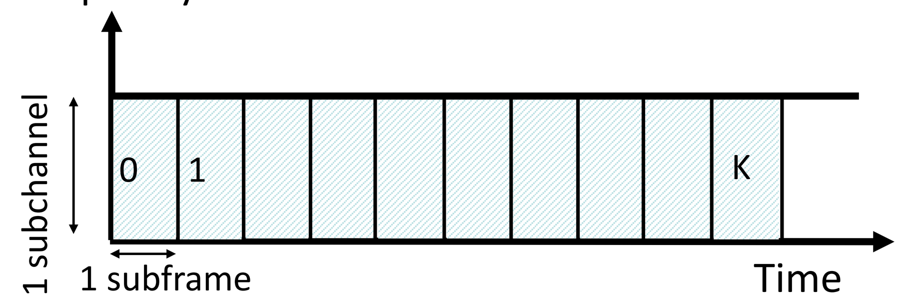
System Model

- N users have to select one of K resources where $K \leq N$.
- If m_j>1 users select a resource only the closer transmitter will be decoded.
- **State Space**
  - Each vehicle i, at time-step t, observe $s_t^i = (a_{t-1}^i, v_t^i)$ with previous action and view-based positional distribution $v_t^i = f(positions, B, R)$ (intuition:[3]). Piggyback positions with periodic safety messages.



v_t^i										
	0.0	0.0	0.33	0.33	0.33	0.0	0.0	0.0	0.0	0.0

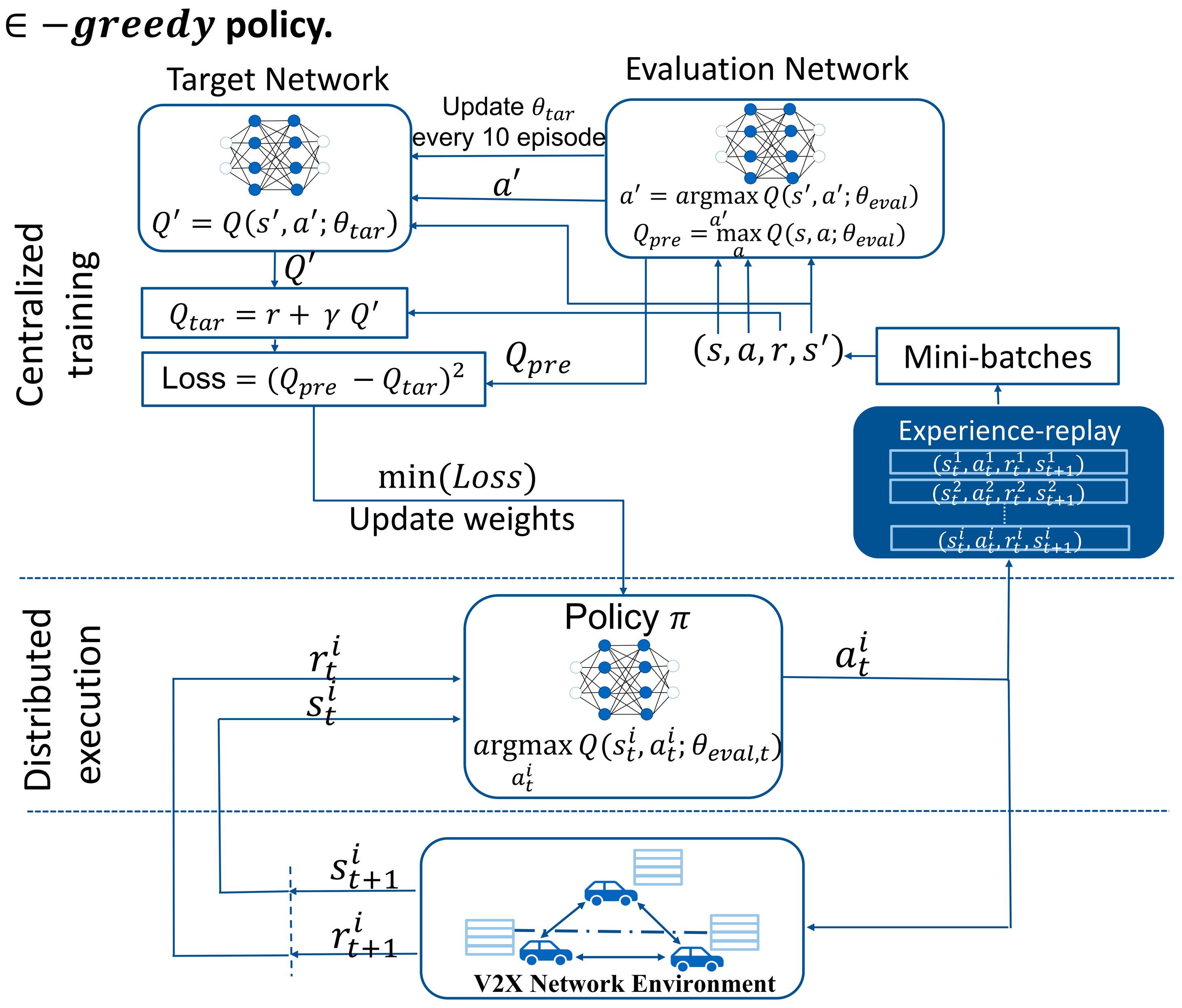
- **Action Space**
  - $a_t^i \in \{1, 2, \dots K\},\ i \in \{1, 2, \dots N\}$ where $K < N$ is the number of resources.



- **Reward Design**
  - Cooperative. N_c^i :=vehicles perform the same action with the user i.

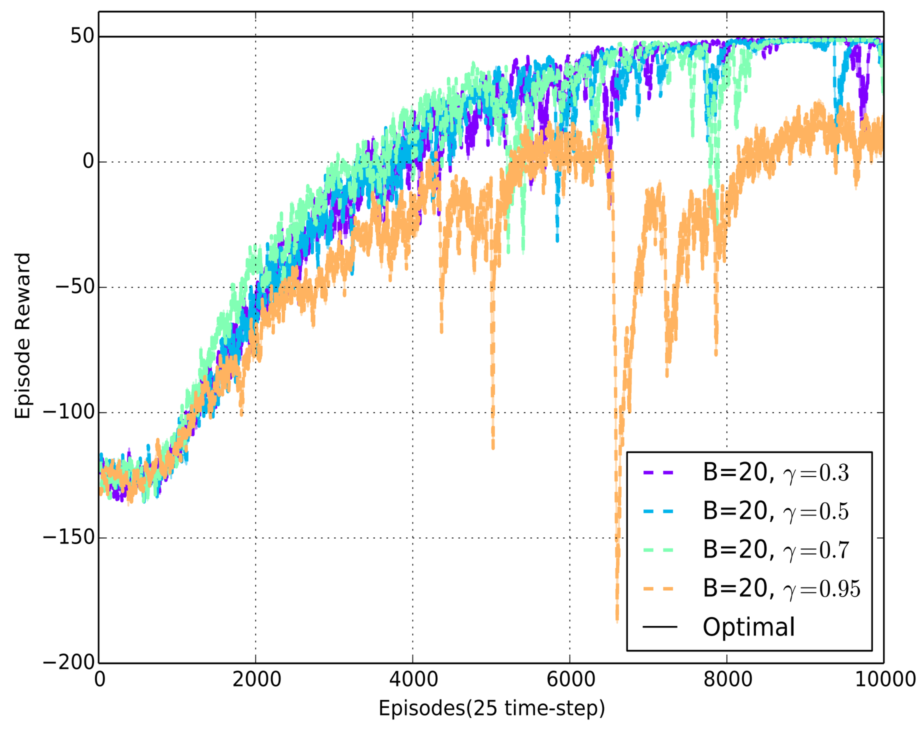
$$r_t^i(a_t^i, s_t^i) = \begin{cases} 1, & N_c^i = 1 \\ \left.\begin{matrix} 0 & \text{if } dist(i,k) \text{ farthest} \\ -N_c^i & \text{else} \end{matrix}\right\} & N_c^i = 2 \\ -N_c^i, & N_c^i > 2 \end{cases} \quad r_t^i(a_t^i, s_t^i) = r_t^i(a_t^i, s_t^i) + \frac{\sum_{i=1}^{n} r_t(i)}{n}$$

- **Goal**: Maximize the discounted return $R_t = \sum_{l=0}^{H} \sum_{i=0}^{N} \gamma^l\, r_{t+l}^i(a_{t+l}^i, s_{t+l}^i)$
- **Double Deep Q Network(DQN) with Long-Term-Short-Memory(LSTM) input layer and $\in -greedy$ policy.**

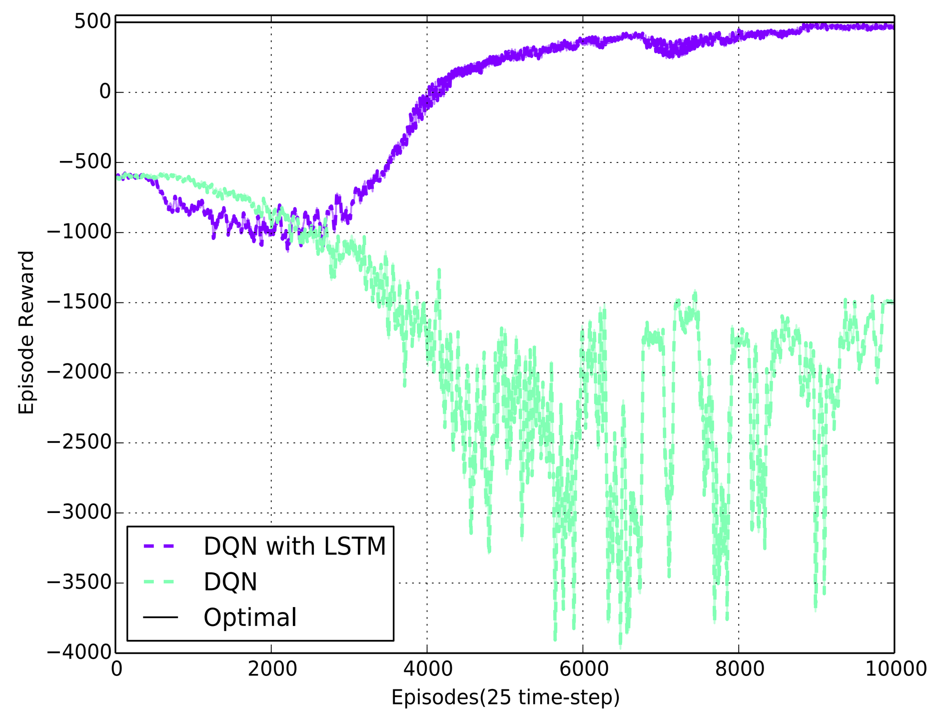


Evaluations

N=4, K=3, 100m highway, velocities={18,36,45,54}kmph



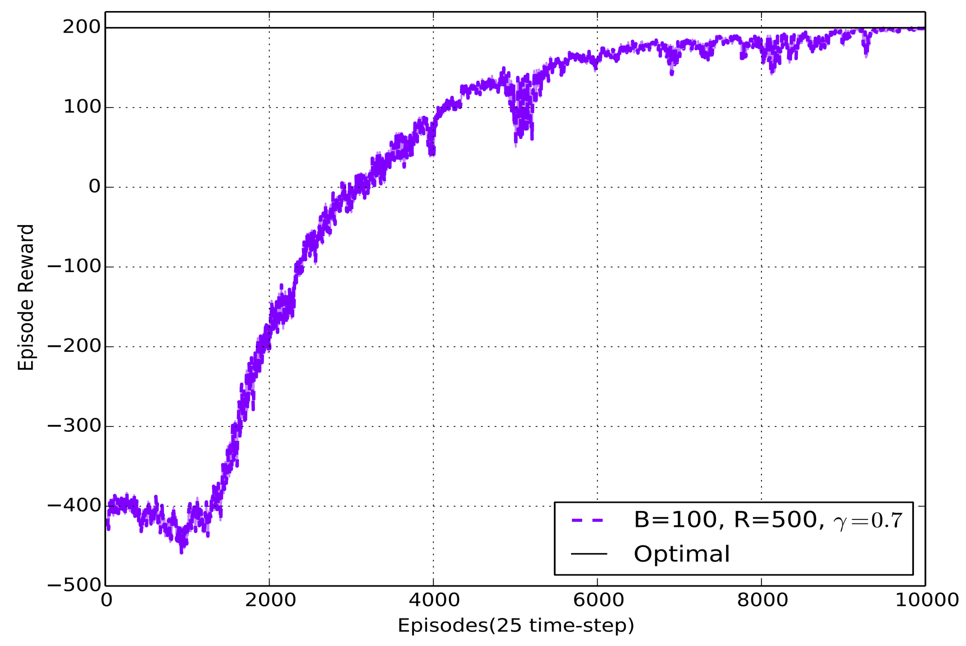
N=20, K=20, 500m highway,~35kmph, SUMO mobility.

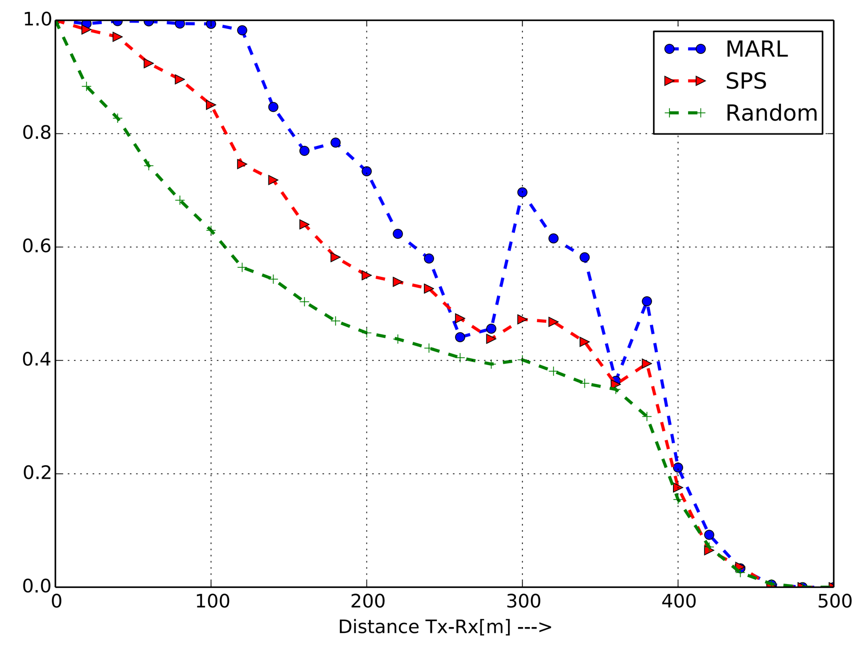


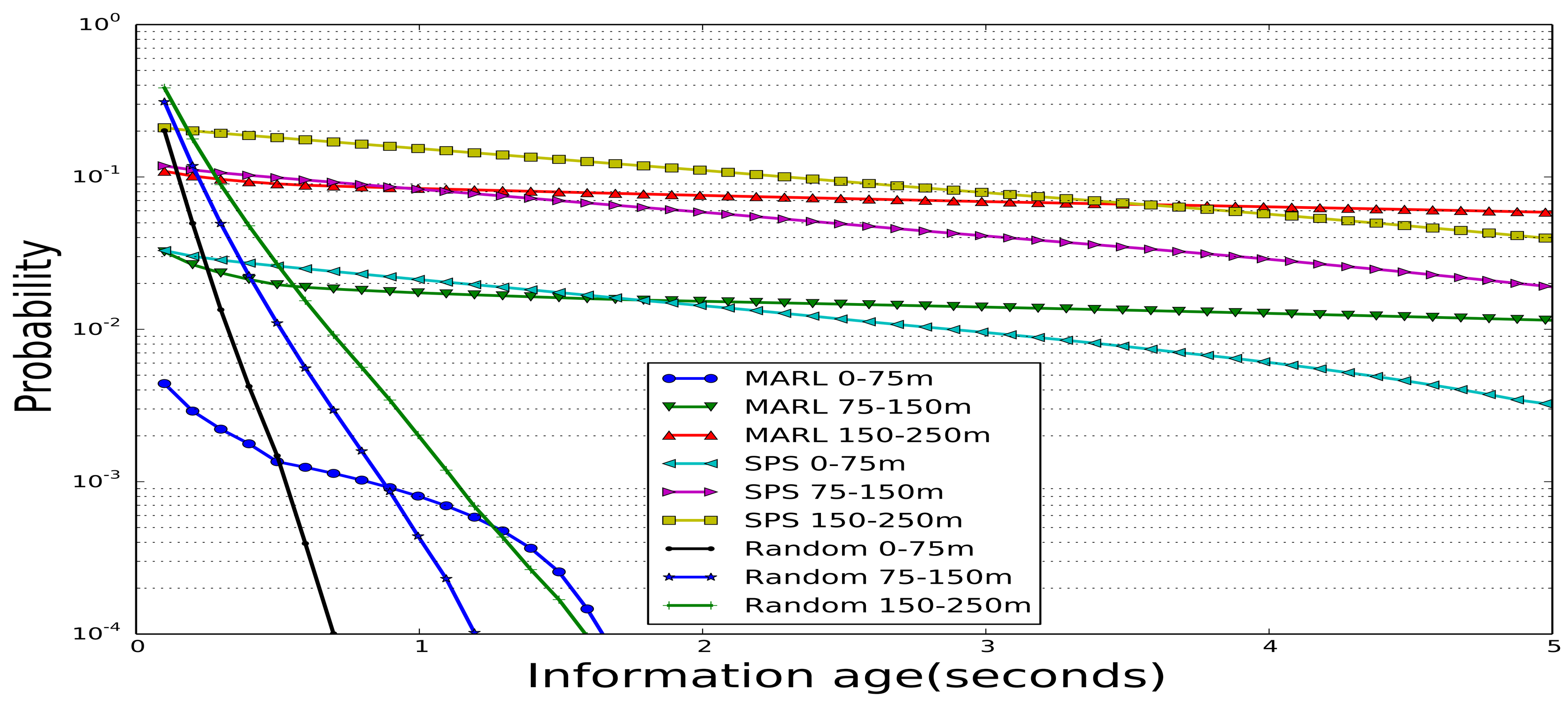
- The proposed approach converges to the desired policy!
- Higher discount factor ($\gamma \geq 0.95$) degrades learning performance.
- LSTM enables learning for the desired policy.

Network Simulations[4]

- Scenario: N=12, K=10, 500m highway, ~35 kmph, SUMO mobility.







- The average packet reception ratio(PRR) over time for the proposed MARL approach is **0.7864** whereas the average PRR for SPS is **0.6913.**

Conclusions

- In congestion, vehicles sacrifice communication with far vehicles in order to provide higher PRR and lower information age(IA) for near vehicles.
- Limitations:
  - Tested only in one direction. Directional resource pools.
- Contributions:
  - Fully distributed resource allocation based on view-based positional distributions of vehicles.
  - Train the policy in simulation and deploy it in real vehicles.
- Future works:
  - Evaluation of different mobility and environment scenarios.
  - Scaling the algorithm for larger scenarios.

References

[1] World Health Organization. Global status report on road safety 2018. Technical report, Genf, Schweiz, 2018.

[2] Nomor Research. Comparison of V2X based on 802.11p, LTE and 5G. White paper, Munich, Germany, April 2019

[3] Foerster, J.N., de Witt, C.A.S., Farquhar, G., Torr, P. H., Boehmer, W., & Whiteson, S. (2018). Multi-agent common knowledge reinforcement learning. arXiv preprint arXiv: 1810.11702.

[4] Realtime Network Simulator(RealNeS), Nomor Research. http://nomor.de/services/simulation/system-level-simulation/