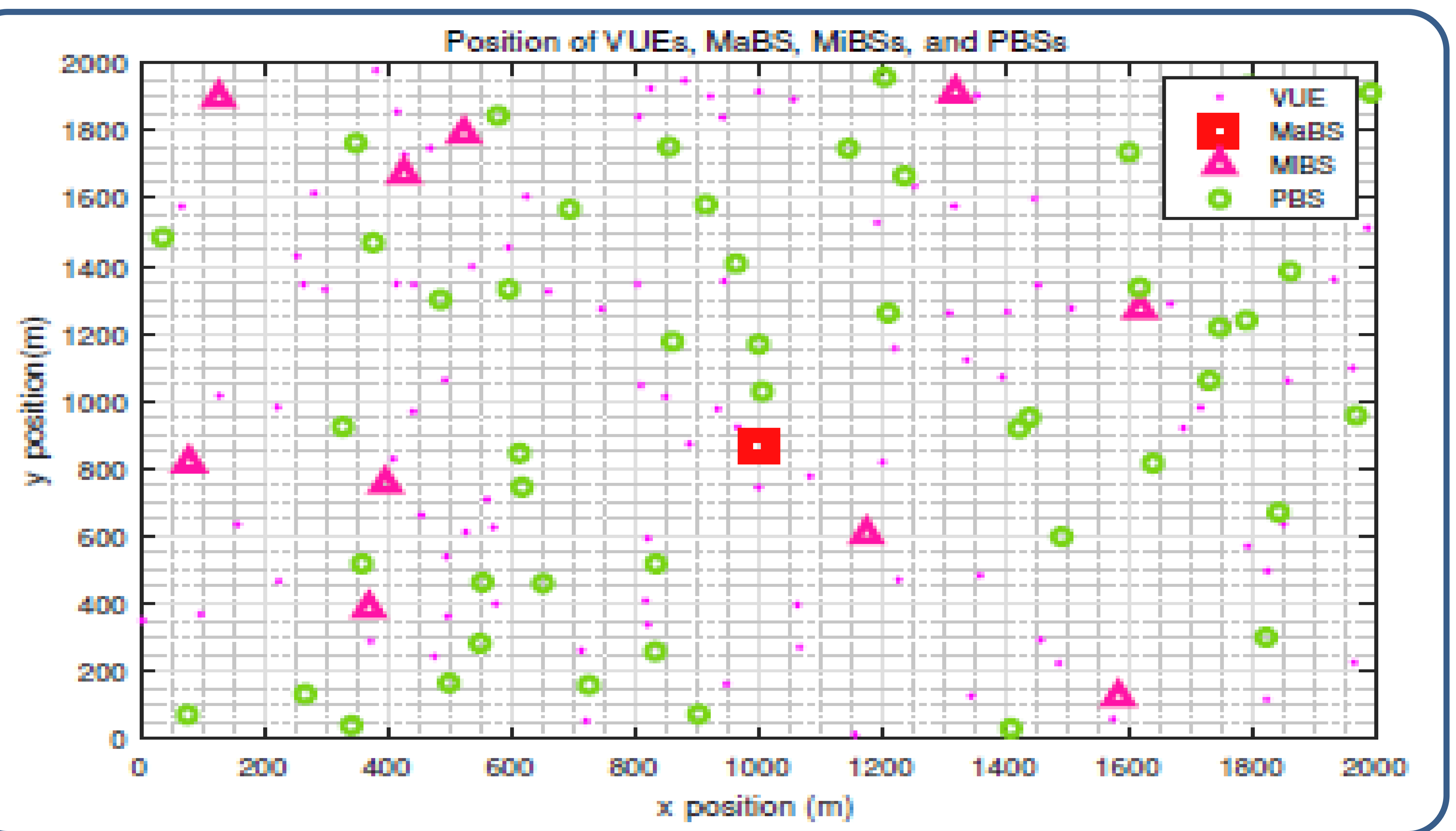
Poster: Multi-Agent Deep Reinforcement Learning for Connected Vehicles

Dohyun Kwon, Soohyun Park, Joongheon Kim (Chung-Ang University, Seoul, Korea)

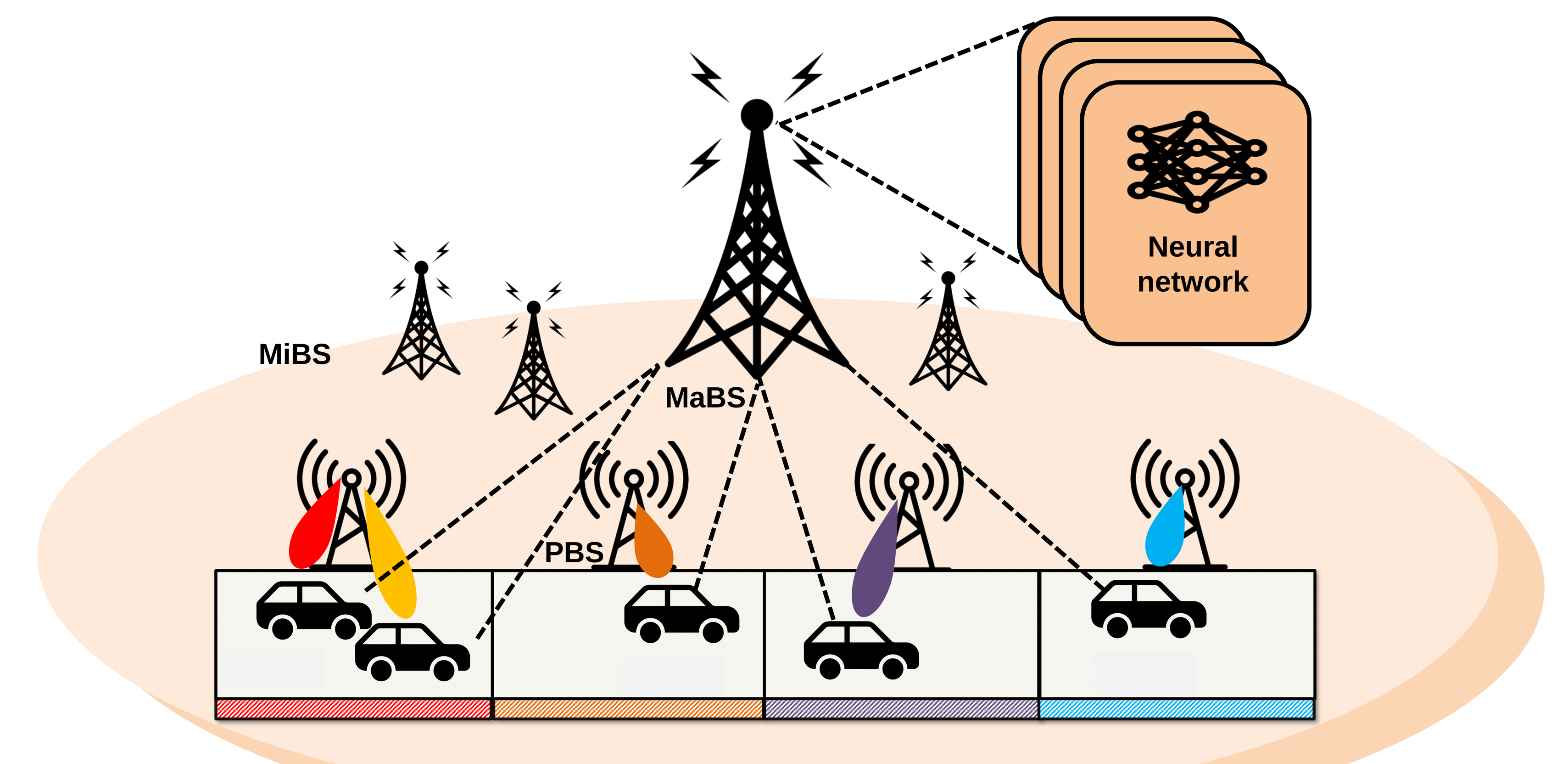
E-mails: kdh1102@cau.ac.kr, shpark.cau@gmail.com, joongheon@cau.ac.kr

Introduction and Backgrounds

Heterogeneous vehicular networks



Connected vehicles in 5G networks



MADDPG for cooperative user association and frequency allocation (UAFA)

State space

(1) Association information between VUEs and BSs
(2) Cumulative downlink traffic volume
(3) Position of VUEs

Action space

(1) For each VUE, what BS to be associated at next time step
(2) What channel to select
(3) If associated BS is PBS, then what direction to set its antenna array to set up the directive link

Reward structure

(1) High reward when VUEs success to orient its antenna array to a PBS and associate with the PBS
(2) Small reward for VUEs when they associate with MaBS/MiBS when they can't associate with PBSs
(3) Penalty for VUEs when they associate with MaBS/MiBS even though they can associate with PBSs

$$Q_{\mu_\theta}(\mathbf{x}, a) = r_{t+1} + \gamma \mathbb{E}_{a\sim\mu(\cdot|\mathbf{x}),\mathbf{x}\sim\mathcal{X}}(r_{t+2} + \cdots + \gamma^{T-2} r_T).$$ ➔ **Action-value function**

$$\nabla_{\theta_i}\mathcal{J} \approx \frac{1}{S}\sum_j \nabla_{\theta_i}\mu_i(o_i^j)\nabla_{a_i}Q_i^{\mu}(\mathbf{x}^j, a_1^j, \cdots, a_N^j|_{a_i=\mu_i(o_i^j)})$$ ➔ **Gradient calculation**

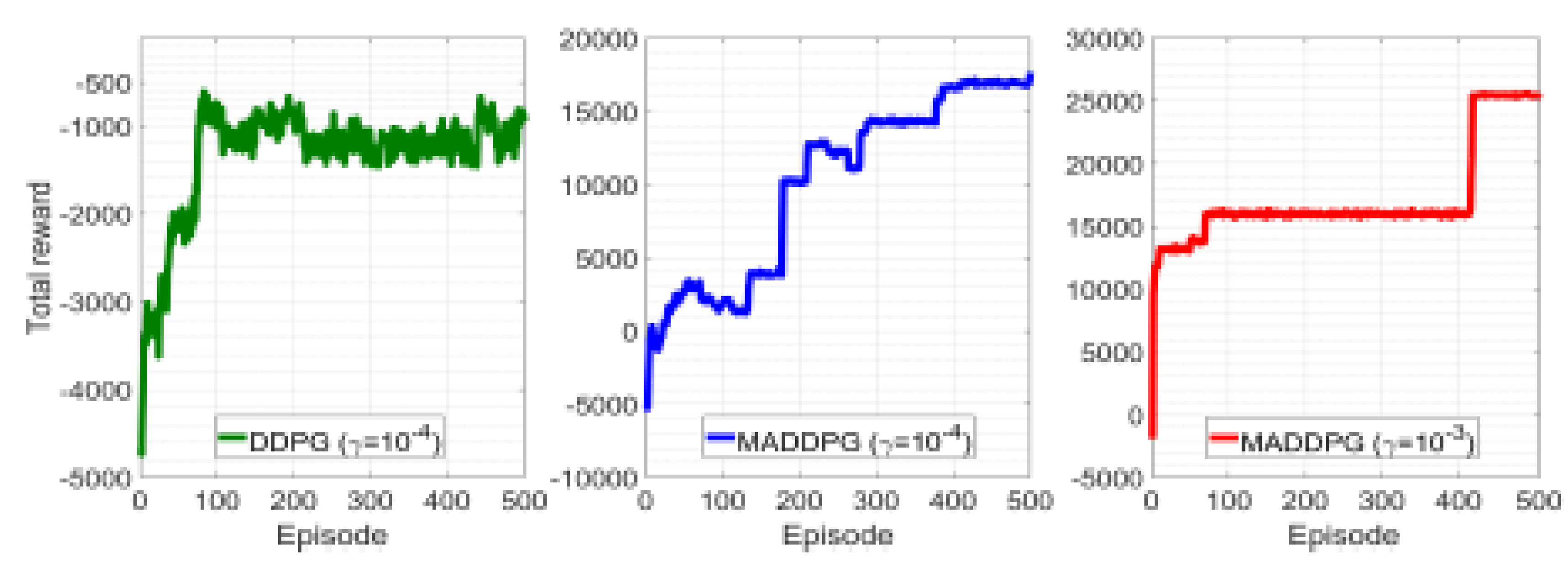
Algorithm description and performance evaluation

Algorithm

Algorithm 1 MADDPG for UAFA in HetVNet

for episode = 1 to E **do**
  Initialize the state of VUEs $\mathbf{x}$ and exploration noise $\mathcal{N}_t$
  for timestep = 1 to T **do**
    Each i–th selects action $a_i = \mu_{\theta_i}(o_i) + \mathcal{N}_t$
    Execute actions a = $(a_1, \cdots, a_N)$
    Observe r and $\mathbf{x}'$ and store $(\mathbf{x}, a, r, \mathbf{x}')$
    for VUE i = 1 to N **do**
      Get S samples $(\mathbf{x}^j, a^j, r^j, \mathbf{x}'^j)$ from D
      Set y^j by Eq. (1)
      Update *critic* by minimizing Eq. (2)
      Update *actor* by Eq. (3)
    Update θ_i' of each VUE

Performance evaluation



The MADDPG algorithm based UAFA solution showed about 25 times superior performance than single agent based model, which is DDPG based one.

The 17th ACM International Conference on Mobile Systems, Applications and Services – Posters (Mobisys 2019 Posters)
June 18, 2019
Seoul, Korea