Xiaoyu LIU, Chi XU, Haibin YU, Peng ZENG, 2022. Multi-agent deep reinforcement learning for end–edge orchestrated resource allocation in industrial wireless networks. *Frontiers of Information Technology & Electronic Engineering,* 23(1):47-60. https://doi.org/10.1631/FITEE.2100331

Multi-agent deep reinforcement learning for end–edge orchestrated resource allocation in industrial wireless networks

Key words: Multi-agent deep reinforcement learning; End–edge orchestrated; Industrial wireless networks; Delay; Energy consumption

Corresponding authors: Chi XU, Haibin YU
E-mail: xuchi@sia.cn; yhb@sia.cn
ORCID: Chi XU, https://orcid.org/0000-0001-7389-5763; Haibin YU, https://orcid.org/0000-0002-1663-2956

Motivation

- For resource-constrained machine-type devices (MTDs) in industrial wireless networks (IWNs), it is a challenge to process computation-intensive and delay-sensitive data locally.
- Within the centralized cloud computing paradigm, offloading data from MTDs to the cloud server triggers significant communication delay, which is intolerable for real-time applications.
- Within the multi-access edge computing paradigm, centralized intelligence triggers significant delay and energy consumption during the collection of global system information, which may be intolerable for real-time applications.
- Distributed MTDs have potential advantages to achieve swarm intelligence. Every MTD acts as an independent agent that can easily observe its local system information, and the cooperation among multiple MTDs can achieve logical resource allocation, and adapt to the dynamic end–edge orchestrated IWNs.

Contributions

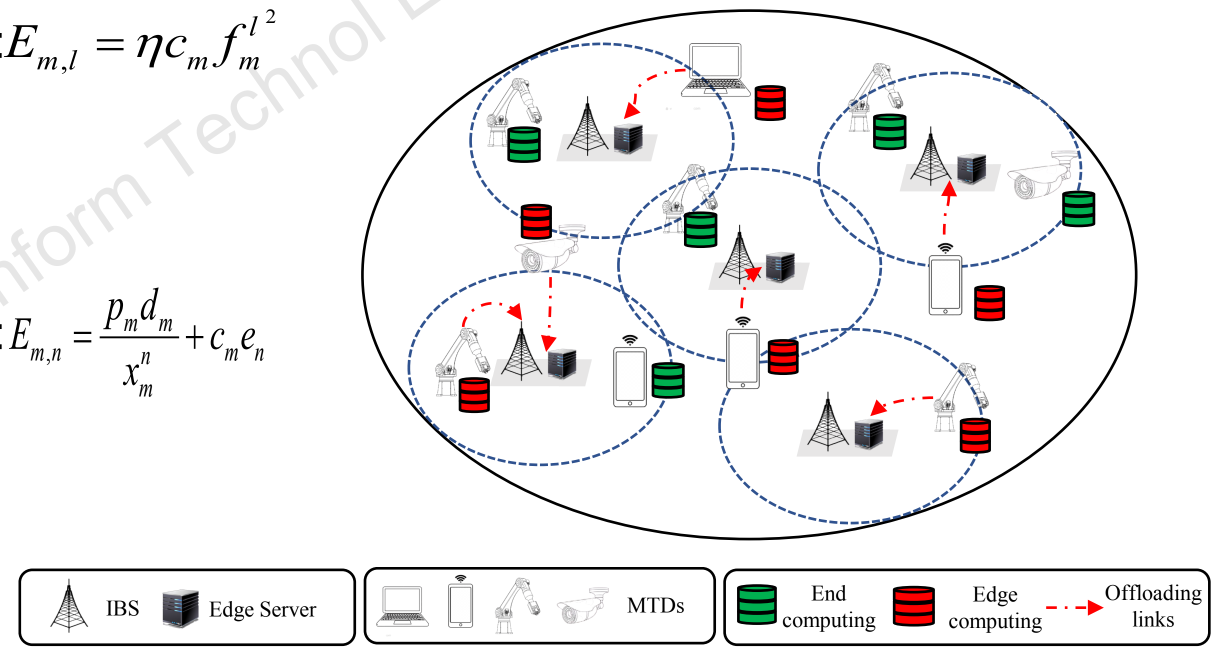
- Considering the diverse constraints on resource allocation, we apply a Markov decision process to formulate the joint optimization of delay and energy consumption, and use multi-agent deep reinforcement learning (MADRL) to learn an effective resource allocation policy for minimizing the system overhead with respect to delay and energy consumption.

- To ensure that the training data are independent and identically distributed while accelerating the learning process of the MADRL-based resource allocation (MADRL-RA) algorithm, we design a weighted experience replay to store and sample experiences categorically.

- To balance the exploration and exploitation of knowledge about IWNs, we propose a step-by-step ε-greedy method to adjust the probabilities of exploration and exploitation dynamically.

System model

- Data of the m^{th} MTD is denoted as $D_m=\{d_m, c_m\}$
- Transmission power of the m^{th} MTD is denoted as $p_m \in \{0, P\}$
- End computing:
  - Delay: $T_{m,l} = c_m \big/ f_m^l$
  - Energy consumption: $E_{m,l} = \eta c_m {f_m^l}^2$
- Edge computing:
  - Delay: $T_{m,n} = \dfrac{d_m}{x_m^n} + \dfrac{c_m}{f_m^n}$
  - Energy consumption: $E_{m,n} = \dfrac{p_m d_m}{x_m^n} + c_m e_n$



Problem formulation

$$
\begin{aligned}
\min_{o_m^j, p_m} \quad & \sum_{m \in \mathcal{M}, j \in \{0, \mathcal{N}\}} o_m^j \left(\omega T_m + (1-\omega) E_m \right) \\
\text{s.t. } & \mathrm{C1}: 0 \le p_m \le P, \\
& \mathrm{C2}: \sum_{m \in \mathcal{M}, n \in \mathcal{N}} o_m^n f_m^n \le F_n, \\
& \mathrm{C3}: 0 \le f_m^n \le F_n, \\
& \mathrm{C4}: \sum_{j=0}^{N} o_m^j = 1, \\
& \mathrm{C5}: o_m^j \in \{0, 1\},
\end{aligned}
$$

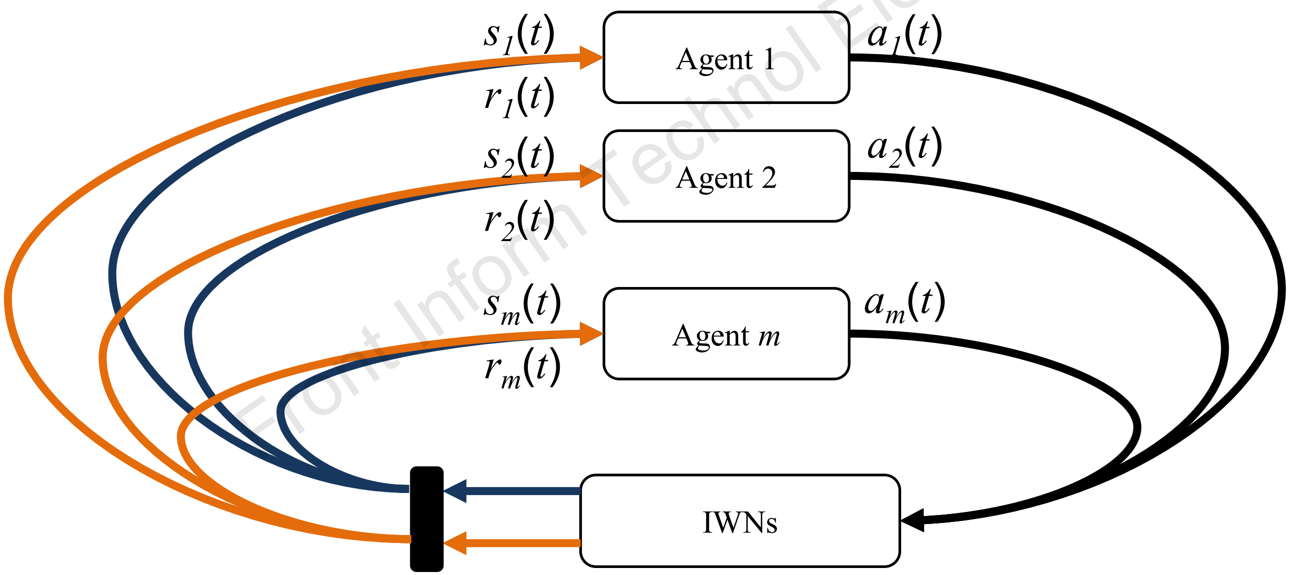
C1 is the transmission power constraint

C2 and C3 are the computation capacity constraints

C4 and C5 are the computing decision constraints

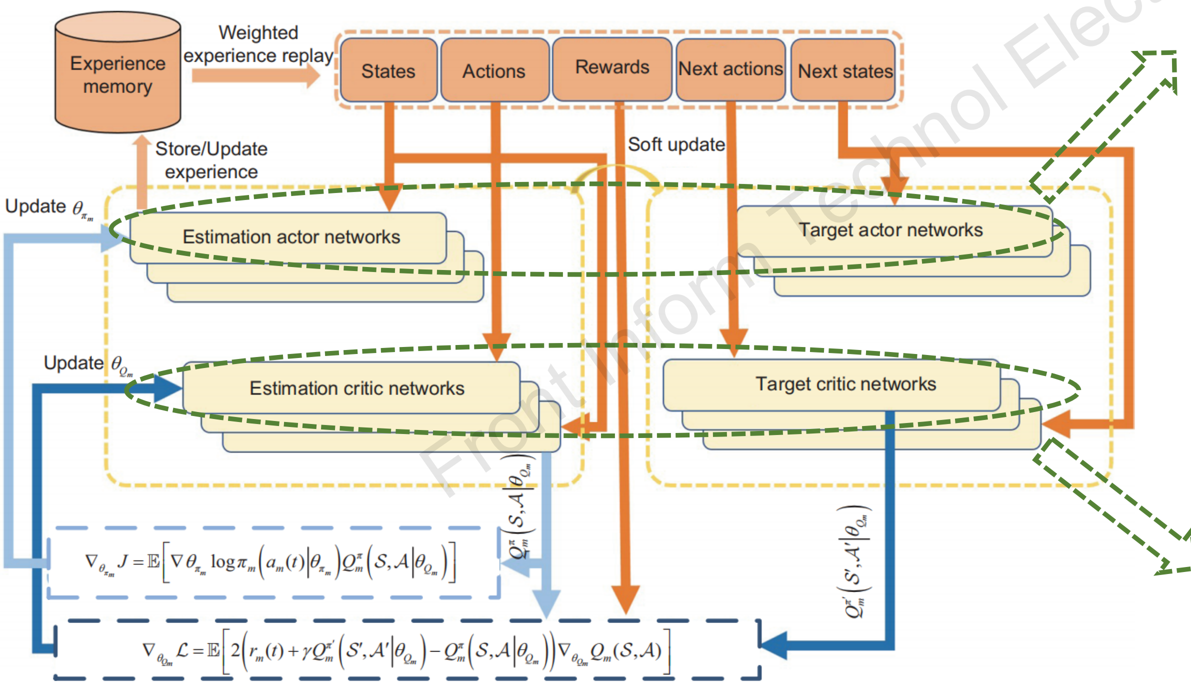
MADRL-RA algorithm

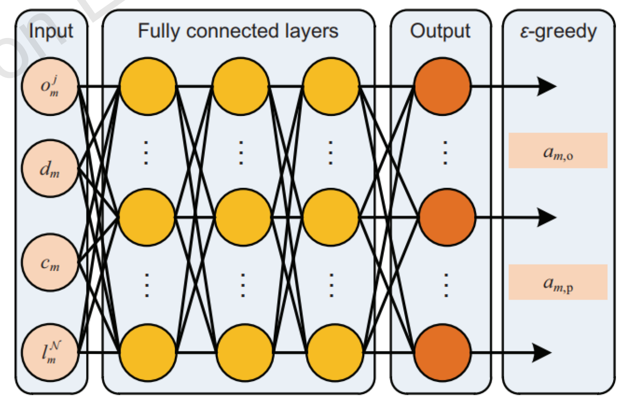
- Every MTD acts as an independent agent that can easily observe its local system information, and the cooperation among multiple MTDs can achieve logical resource allocation, and adapt to the dynamic end–edge orchestrated IWNs.



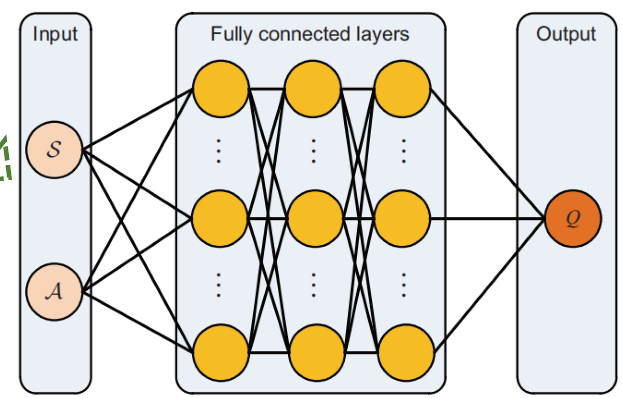
MADRL-RA algorithm

- For an independent agent, the actor is used to generate action, and the critic is used to guide the actor in generating a better action. The experience memory is used to store experiences for training the actors and critics.





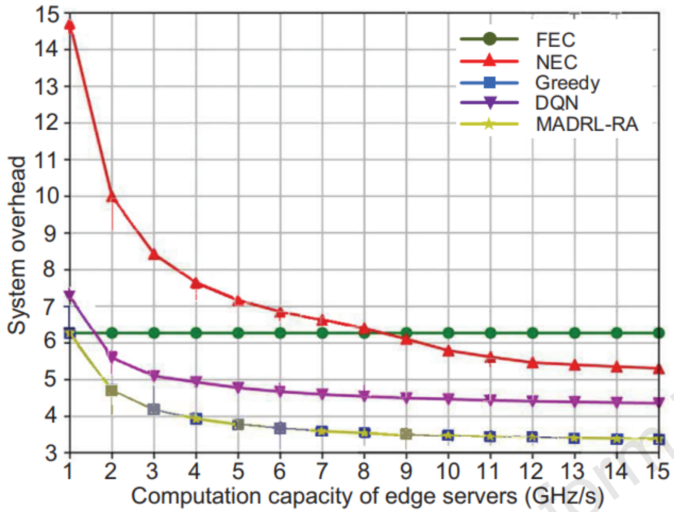
Actor networks

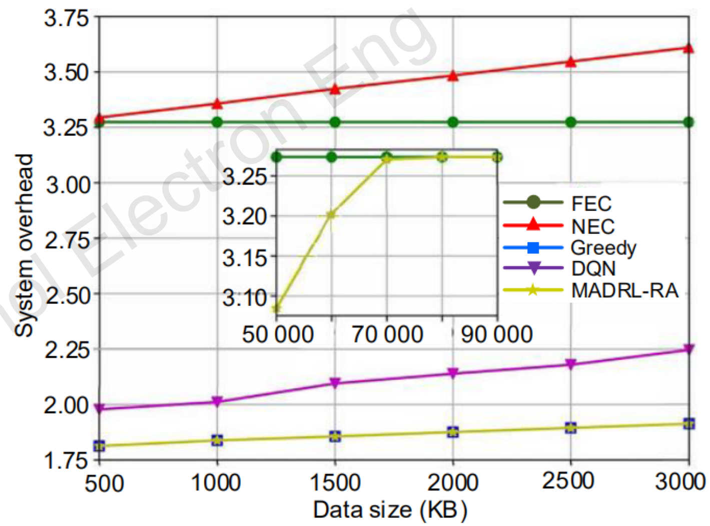


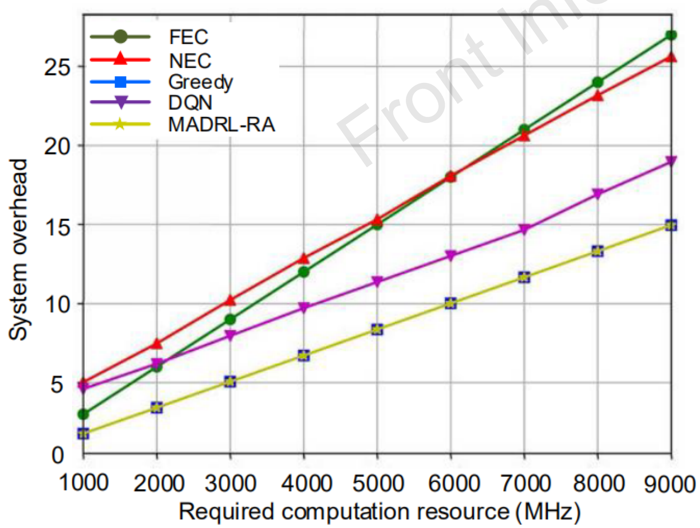
Critic networks

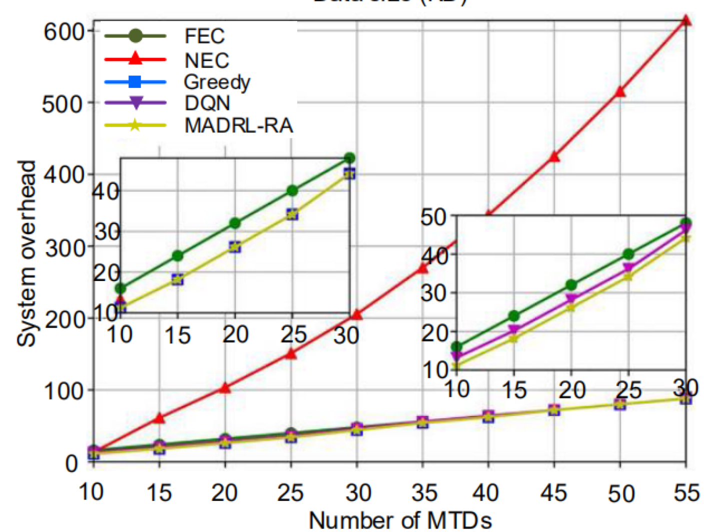
Major results

- MADRL-RA can adapt to the changes of edge server computing capacity, data size, required computing resources, and the number of devices, learn the optimal resource optimization strategy, and achieve the minimum system overhead.



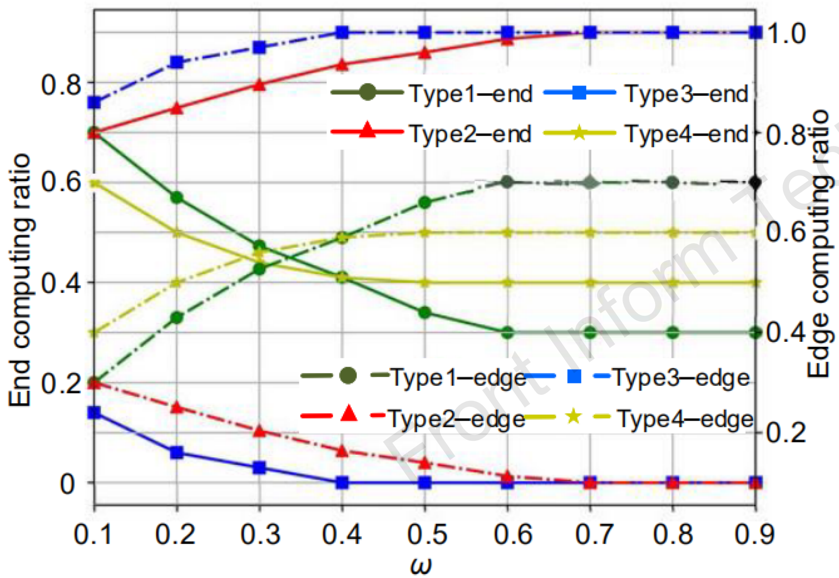


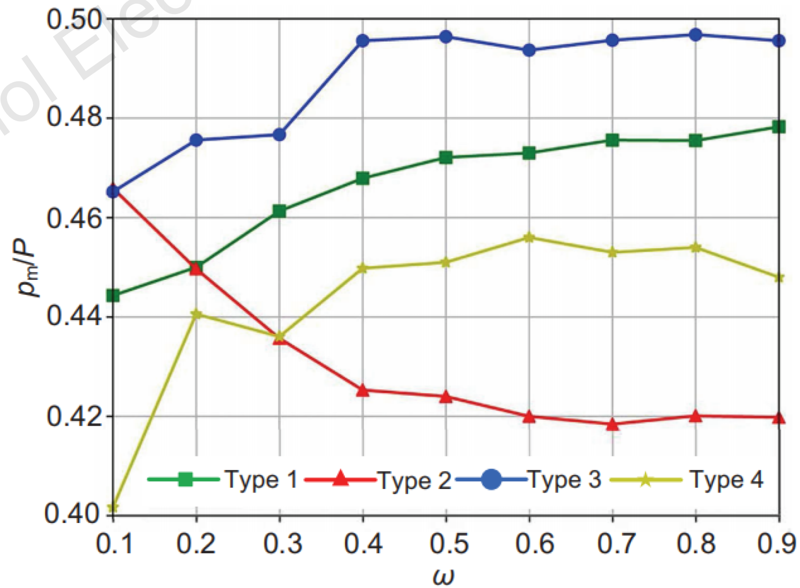




Major results

- By setting different weight factors, MADRL-RA can optimize resource allocation to satisfy heterogeneous industrial applications in terms of different delay or energy consumption.





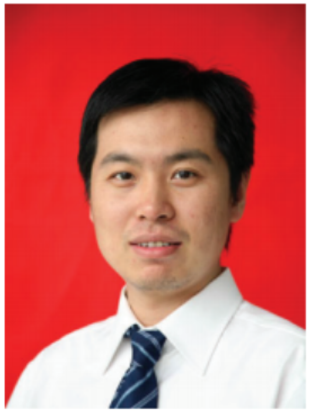
Conclusions

- We regarded distributed MTDs as multiple self-learning agents to deal well with dynamic and time-varying end–edge orchestrated IWNs.
- We proposed the MADRL-RA algorithm to learn an end–edge orchestrated resource allocation policy for minimizing system overhead with respect to delay and energy consumption.
- Compared with FEC, NEC, and DQN, MADRL-RA can learn the effective resource allocation policy and adapt to changes in the computation capacity of edge servers, data size, required computation resource, and the number of MTDs.
- By setting different weight factors, we can optimize resource allocation to satisfy heterogeneous industrial applications in terms of different delay or energy consumption.

Xiaoyu LIU is currently pursuing the PhD degree in control theory and control engineering with University of Chinese Academy of Sciences, Beijing, China. His research interests include industrial wireless networks and deep reinforcement learning.

Chi XU received the PhD degree from University of Chinese Academy of Sciences in 2017. He is currently an Associate Professor with Shenyang Institute of Automation, Chinese Academy of Sciences. He is a voting member of the IEEE 1918.1 working group for tactile Internet and a member of the IEEE 1932.1 working group for licensed/unlicensed spectrum interoperability in wireless mobile networks. He serves as a standardization delegate for 3GPP TSG RAN. His research interests include industrial wireless networks, 5G URLLC, and tactile Internet.

Haibin YU received the PhD degree from Northeastern University, China, in 1997. He has been a Professor with Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, since 1997. He has published two books, authored or co-authored more than 200 papers, and held more than 50 patents. He and his research team have proposed the WIA-PA and WIA-FA standards, which are specified as IEC 62601 and IEC 62948, respectively. He was elected as an ISA Fellow for his contributions in fieldbus technologies in 2011. He serves as the Vice Chair of Chinese Association of Automation and the Chair of China National Technical Committee for Industrial Process Measurement Control and Automation Standardization. His research interests include wireless sensor networks, industrial communication and networked control, industrial automation, and intelligent manufacturing.



Peng ZENG received the PhD degree from Shenyang Institute of Automation, Chinese Academy of Sciences, China, in 2005. He has been a Professor with Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, since 2007. He is an expert member of the IEC/TC65/WG16, a member of the standards committee of SP100, and a member of the Wireless WG of FieldBus Foundation. He also serves as the Chair of Edge Computing Technical Committee, Chinese Association of Automation. His research interests include wireless sensor networks and industrial wireless communications.