Xiaojun BAI, Yang ZHANG, Haixing WU, Yuting WANG, Shunfu JIN, 2024. A cloud–edge–device collaborative offloading scheme with heterogeneous tasks and its performance evaluation. *Frontiers of Information Technology & Electronic Engineering*, 25(5):664-684. https://doi.org/10.1631/FITEE.2300128

A cloud–edge–device collaborative offloading scheme with heterogeneous tasks and its performance evaluation

Key words: Edge computing; Offloading scheme; Cloud–edge–device collaboration; Markov chain; Cost function

Corresponding author: Shunfu JIN
E-mail: jsf@ysu.edu.cn
ORCID: https://orcid.org/0000-0002-5845-5601

Motivation

1. Edge computing proposed as a new computing paradigm extends computational resources to the edge of the networks and enables data to be processed and analyzed locally in proximity to end-users.
2. The integrated cloud–edge–device collaboration architecture is expected largely to meet the quality of service (QoS) demands of delay-sensitive Internet of Things (IoT) applications and to reduce network bandwidth usage.
3. The division of tasks in cloud and edge computing helps prioritize and allocate resources effectively.
4. Markov chains facilitate the analysis of long-term behaviors and steady-state properties of systems and can capture the random behaviors of different types of tasks.

Method

A task offloading scheme with heterogeneous tasks in a stable cloud–edge–device collaborative environment

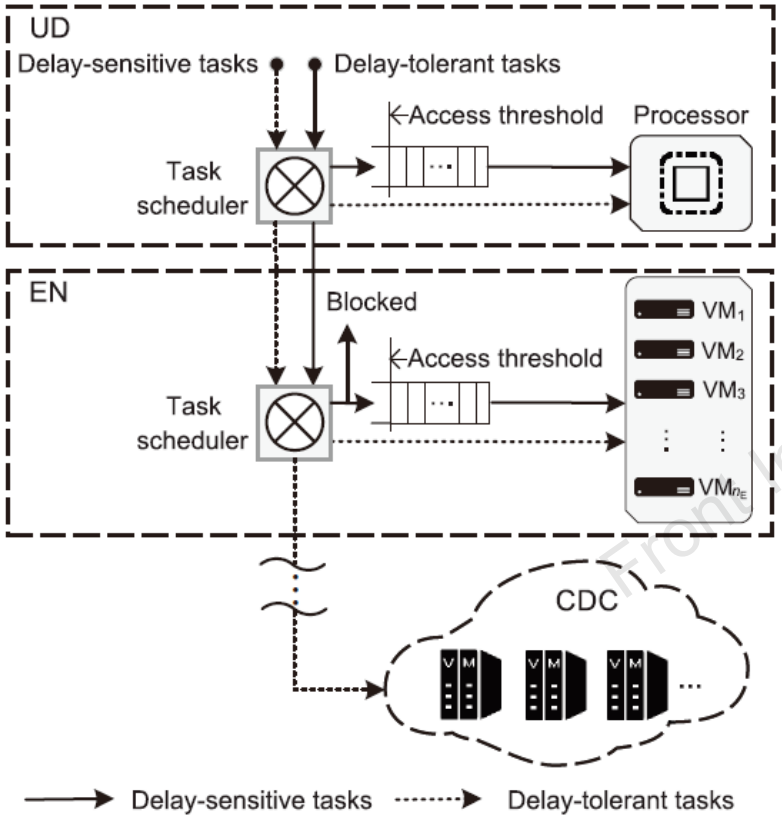

Fig. 1 Working principle of the task offloading scheme (CDC: cloud data center; EN: edge network; UD: user device)

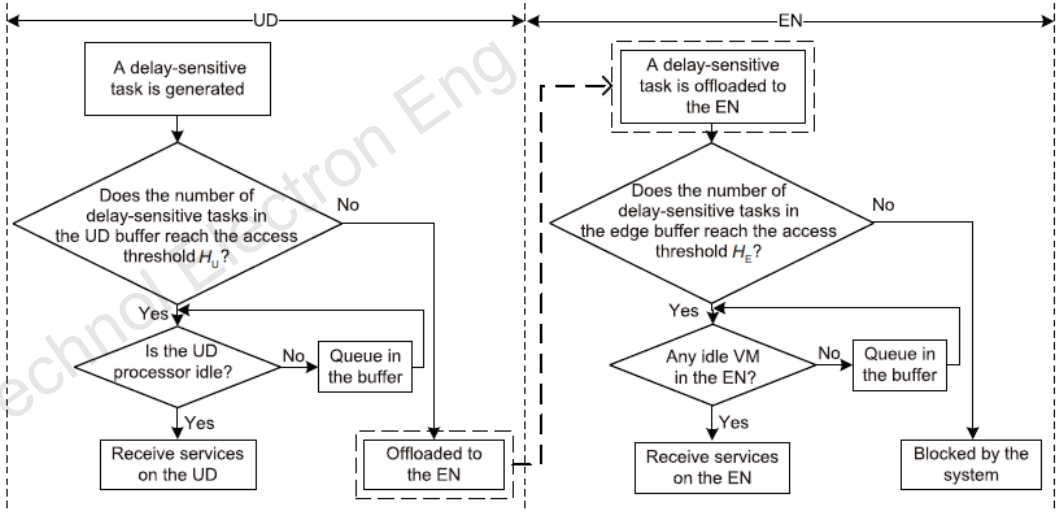

Fig. 2 Workflow of delay-sensitive tasks under the proposed scheme (EN: edge network; UD: user device; VM: virtual machine)

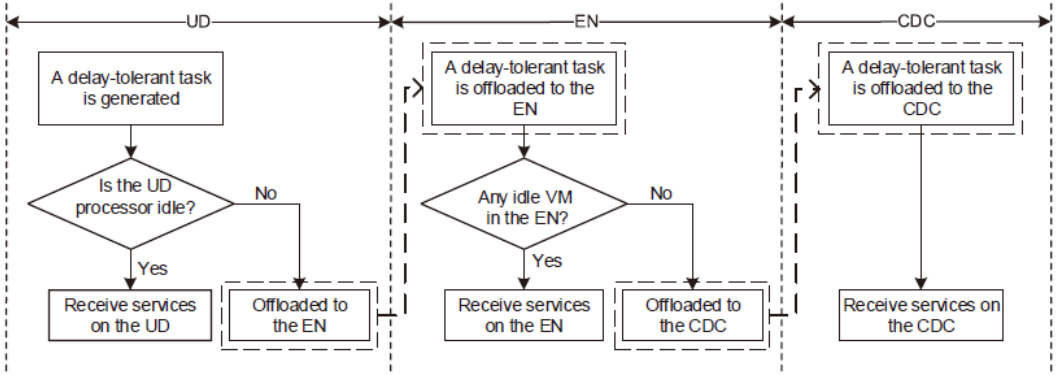

Fig. 3 Workflow of delay-tolerant tasks under the proposed scheme (CDC: cloud data center; EN: edge network; UD: user device; VM: virtual machine)

Method

A four-dimensional continuous-time Markov chain $\{(X(t), Y(t), Z(t), G(t)), t \geq 0\}$ is established as the system model to describe the proposed task offloading scheme and further optimize the system performance.

The state space of this Markov chain is

$$\Omega = \Big\{ (i,j,k,r) \mid i \in \{0,1,\ldots,n_{\mathrm{E}}\}, \quad j \in \{i, i+1,\ldots,n_{\mathrm{E}}+H_{\mathrm{E}}\}, \quad k \in \{0,1\}, r \in \{k, k+1,\ldots,1+H_{\mathrm{U}}\} \Big\}. \tag{1}$$

The generator of the four-dimensional CTMC $\{(X(t), Y(t), Z(t), G(t)), t \geq 0\}$ is

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_{0,0} & \boldsymbol{Q}_{0,1} & & & & \\ \boldsymbol{Q}_{1,0} & \boldsymbol{Q}_{1,1} & \boldsymbol{Q}_{1,2} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \boldsymbol{Q}_{n_{\mathrm{E}}-1,n_{\mathrm{E}}-2} & \boldsymbol{Q}_{n_{\mathrm{E}}-1,n_{\mathrm{E}}-1} & \boldsymbol{Q}_{n_{\mathrm{E}}-1,n_{\mathrm{E}}} & \\ & & & \boldsymbol{Q}_{n_{\mathrm{E}},n_{\mathrm{E}}-1} & \boldsymbol{Q}_{n_{\mathrm{E}},n_{\mathrm{E}}} \end{bmatrix}.$$

Method

Numerical experiments are conducted and analyzed to evaluate the system performance, and numerical simulations are presented to evaluate and validate the effectiveness of the proposed task offloading scheme.

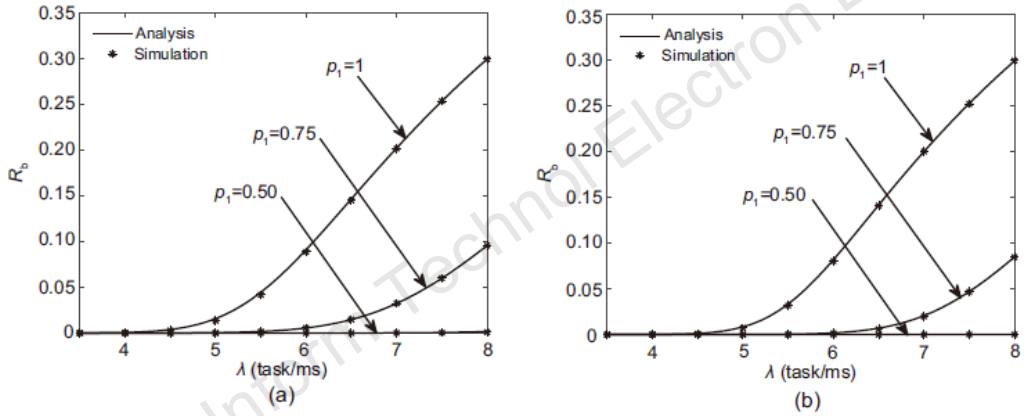

Fig. 4 Change trend for the blocking rate R_{b} of delay-sensitive tasks: (a) $H_{\mathrm{E}} = 15$; (b) $H_{\mathrm{E}} = 20$

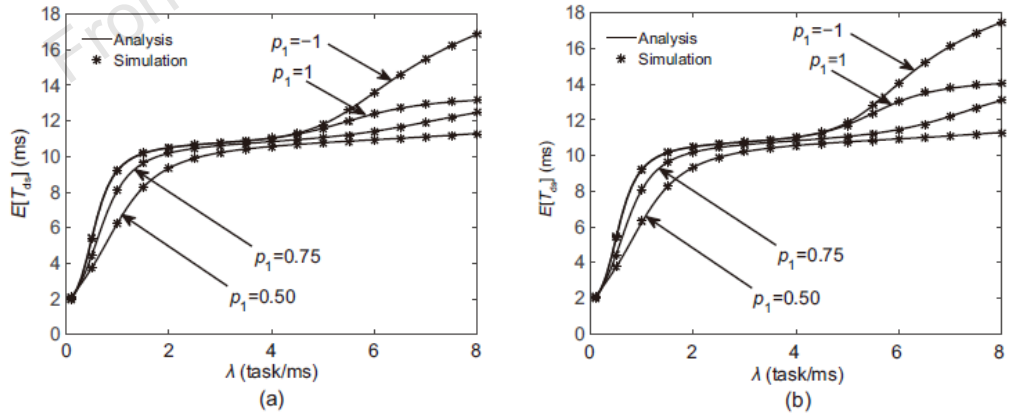

Fig. 5 Change trend for the average delay $E[T_{\mathrm{ds}}]$ of delay-sensitive tasks: (a) $H_{\mathrm{E}} = 15$; (b) $H_{\mathrm{E}} = 20$

Method

To obtain the optimal access threshold H_E in the edge network (EN) buffer, a system cost function $F(H_E)$ based on the analytical results of the system model is constructed.

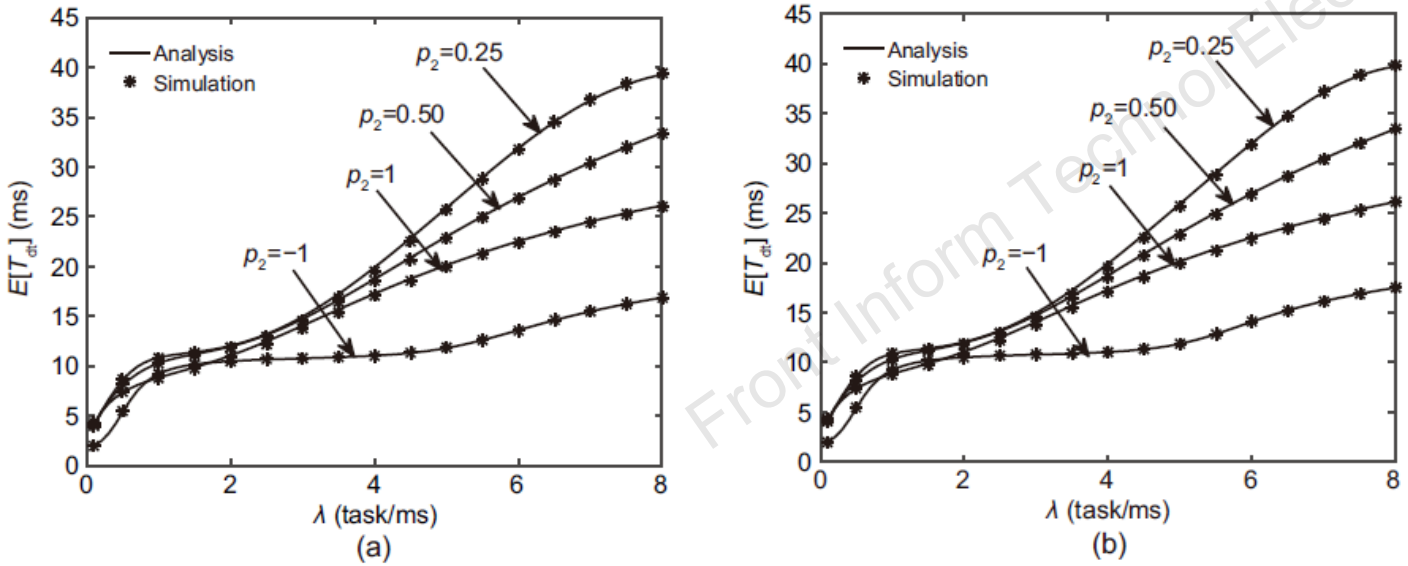

Fig. 6 Change trend for the average delay $E[T_{dt}]$ of delay-tolerant tasks: (a) $H_E = 15$; (b) $H_E = 20$

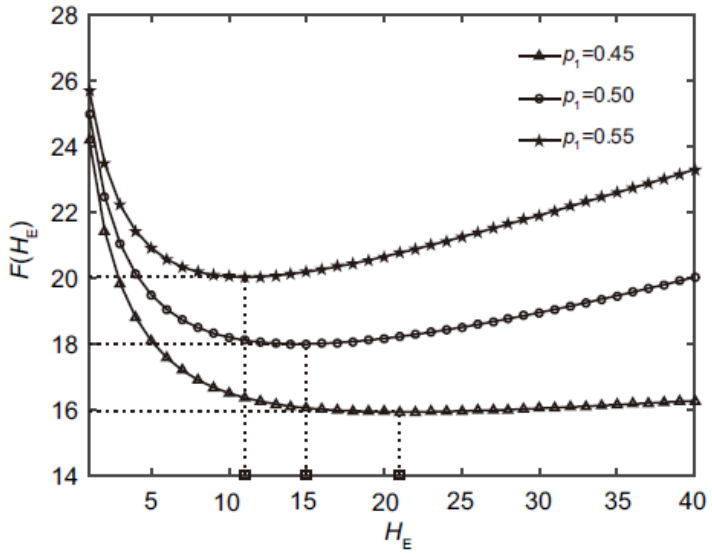

Fig. 7 System cost function $F(H_E)$ versus access threshold H_E in the edge network buffer

Conclusions

1. We proposed a novel task offloading scheme in the cloud–edge–device collaborative edge computing environment to meet differentiated requirements of heterogeneous tasks.
2. By establishing a four-dimensional continuous-time Markov chain as the system model and using the Gauss–Seidel method, we evaluated the long-term performance of the system.
3. We constructed the system cost function and obtained the optimal access threshold in the EN buffer with different proportions of delay-sensitive tasks.

Xiaojun BAI is currently pursuing the Ph.D. degree in computer science and technology at Yanshan University. His research interests include resource management in cloud computing and edge computing, and multi-objective optimization.

Yang ZHANG is a postgraduate at the School of Information Science and Engineering, Yanshan University, Qinhuangdao, China. Her research interests include queuing systems and performance evaluation of edge computing system.

Haixing WU is currently pursuing the Ph.D. degree in computer science and technology at Yanshan University. Her research interests include performance evaluation and resource allocation of cloud computing and mobile edge computing.

Yuting WANG is currently pursuing the Ph.D. degree with the School of Information Science and Engineering, Yanshan University. Her research interests include mobile–edge computing, performance evaluation, and queuing game.

Shunfu JIN is currently a professor with the School of Information Science and Engineering, Yanshan University. Her research interests include artificial intelligence, stochastic modeling for queuing systems, and performance evaluation for communication networks.