SkyMarket: Promoting Cooperation in Competitive Sky Computing Marketplace

Tianchi Su, Qian Ma, Lin Gao, and Xu Chen

Appendix A
Proof of Lemma 1

Since the payoff function of users is concave, users gain the maximum payoff when d_k satisfies the first-order condition. This occurs when:

$$R'_{user}(d_k^*) = \frac{vq_k}{1+q_k d_k^*} - p_k = 0,$$

leading to the optimal demand:

$$d_k^* = \frac{v}{p_k} - \frac{1}{q_k}.$$

Users only choose the SP with the largest φ, any other SPs with smaller φ will not receive demand from the users.

Appendix B
Proof of Theorem 1

We first prove that $p_k^* = c_k$ is a Nash equilibrium when $\eta_{max} = \eta_{sec}$. Under the condition that $p_k^* = c_k, \forall k \in \mathcal{M}$, all SPs obtain zero profit. If any SP $k \in \mathcal{M}$ attempts to set its price $p_k > c_k$, it can not gain a positive profit. This is due to the presence of at least one SP with a higher quality-price ratio, and users would not choose SP k. Next, we demonstrate that no alternative price vector can constitute an equilibrium. If SP k sets its price over cost, there is always at least one SP that could set a lower price, thereby surpassing k_{max}'s quality-price ratio, and SP k's profit would be zero.

When $\eta_{max} > \eta_{sec}$, if $p_{max} < \frac{q_{max}}{q_{sec}} c_{sec}$, the SP with the second-largest quality-cost ratio has no chance to attract users from SP k_{max}. To maximize its profit, SP k_{max} will set its price that $p_{max} = \frac{q_{max}}{q_{sec}} c_{sec} - \varepsilon$. In this way, other SPs would set their price as cost, since no SP could get profit with lower or higher price. If SP k_{max} try to increase its price that $p_{max} > \frac{q_{max}}{q_{sec}} c_{sec}$, SP k_{sec} could set a relatively low price to make its quality-price ratio larger than that of k_{max}'s, in this case, the profit of k_{max} will be zero. When there is only one SP in set $\mathcal{M}$, the optimal price of k_{max} is $\sqrt{q_k c_k v}$. To prove the optimal price, we first substitute $d_k^*(\boldsymbol{p})$ into $R_k(p_k, \boldsymbol{p}_{-k})$,

$$R_k(p_k, \boldsymbol{p}_{-k}) = v - \frac{p_k}{q_k} - \frac{c_k v}{p_k} + \frac{c_k}{q_k},$$

$$\frac{dR_k}{dp_k} = \frac{c_k v}{p_k^2} - \frac{1}{q_k},$$

$$\frac{d^2 R_k}{dp_k^2} = -\frac{c_k v}{p_k^3} \le 0.$$

This confirms the strict concavity of the optimization problem for the SP, ensuring a unique solution. Setting the first derivative equal to zero yields the optimal price:

$$\frac{dR_k}{dp_k} = \frac{c_k v}{p_k^{*2}} - \frac{1}{q_k} = 0,$$

$$p_k^* = \sqrt{q_k c_k v}.$$

This completes our proof.

Appendix C
Proof of Lemma 2

The existence of MPE has been proved in [1].

Appendix D
Proof of Theorem 2

In Stage II, we calculate the profit of coalition $\mathcal{B}, \mathcal{B} \in \mathbf{\Omega}$ as follows:

$$R_{\mathcal{B}}(\mathbf{\Omega}) = \begin{cases} \left(\sqrt{v} - \sqrt{\frac{c_{\mathcal{B}}}{q_{\mathcal{B}}}}\right)^2, & \text{if } \frac{q_{\mathcal{B}}}{q_{sec}} c_{sec} \ge \sqrt{q_{\mathcal{B}} c_{\mathcal{B}} v}, \\ & \eta_{\mathcal{B}} > \eta_{sec}, \\ v - \frac{c_{sec}}{q_{sec}} - \frac{c_{\mathcal{B}}}{q_{\mathcal{B}}}\left(\frac{v q_{sec}}{c_{sec}} - 1\right), & \text{if } \frac{q_{\mathcal{B}}}{q_{sec}} c_{sec} < \sqrt{q_{\mathcal{B}} c_{\mathcal{B}} v}, \\ & \eta_{\mathcal{B}} > \eta_{sec}, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

From Eq. (1), we can see when $\mathcal{B}$ wins the price competition game, as the quality-cost ratio $\eta_{\mathcal{B}} = \frac{q_{\mathcal{B}}}{c_{\mathcal{B}}}$ increases, the profit of coalition $\mathcal{B}$ increases. Suppose a CSP (or coalition) $i \in \mathbf{\Omega}$ joins the coalition $\mathcal{B}$, forming a new coalition$\mathcal{B} \cup i$, the quality-cost ratio $\eta_{\mathcal{B}\cup i}$ is given by:

$$\eta_{\mathcal{B}\cup i} = \frac{(q_{\mathcal{B}} + q_i)^2}{q_{\mathcal{B}} c_{\mathcal{B}} + q_i c_i}.$$

When

$$\eta_{\mathcal{B}\cup i} - \eta_{\mathcal{B}} = \frac{(q_{\mathcal{B}} + q_i)^2}{q_{\mathcal{B}} c_{\mathcal{B}} + q_i c_i} - \frac{q_{\mathcal{B}}}{c_{\mathcal{B}}} > 0,$$

$$\begin{aligned} c_{\mathcal{B}}(q_{\mathcal{B}} + q_i)^2 &> q_{\mathcal{B}}(q_{\mathcal{B}} c_{\mathcal{B}} + q_i c_i), \\ c_{\mathcal{B}} q_{\mathcal{B}}^2 + 2c_{\mathcal{B}} q_i q_{\mathcal{B}} + c_{\mathcal{B}} q_i^2 &> q_{\mathcal{B}}^2 c_{\mathcal{B}} + q_{\mathcal{B}} q_i c_i, \\ c_{\mathcal{B}} q_i^2 + q_{\mathcal{B}} q_i (2c_{\mathcal{B}} - c_i) &> 0, \\ c_{\mathcal{B}} q_i + q_{\mathcal{B}}(2c_{\mathcal{B}} - c_i) &> 0, \\ q_i &> q_{\mathcal{B}}\left(\frac{c_i}{c_{\mathcal{B}}} - 2\right). \end{aligned}$$

Tianchi Su and Qian Ma are with the School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, Guangdong 518107, China. E-mail: sutch@mail2.sysu.edu.cn, maqian25@mail.sysu.edu.cn.

Lin Gao is with the School of Electronics and Information Engineering, Harbin Institute of Technology, Shenzhen, China. E-mail: gaol@hit.edu.cn.

Xu Chen is with the School of Computer Science and Engineering, Sun Yat-sen University, China. E-mail: chenxu35@mail.sysu.edu.cn.

It denotes that when $c_i \le 2c_{\mathcal{B}}$, or $c_i > 2c_{\mathcal{B}}$ and $q_i > q_{\mathcal{B}}\left(\frac{c_i}{c_{\mathcal{B}}} - 2\right)$ the inequality always holds, which means the joined coalition i can increase the profit of coalition $\mathcal{B}$.

First, we assume grand coalition $\mathcal{G}$ is inefficient when $c_{max} \le 2c_{min}$, i.e. for any coalition structure $\mathbf{\Omega}$, $R_{\mathcal{G}} < \sum_{\mathcal{B}\in\mathbf{\Omega}} R_{\mathcal{B}}(\mathbf{\Omega})$, in which $R_{\mathcal{G}}$ is the profit of grand coalition. There is only one coalition that wins the price competition can gain positive profit, and we denote the profit of the winner as $R_{win}(\mathbf{\Omega}) = \sum_{\mathcal{B}\in\mathbf{\Omega}} R_{\mathcal{B}}(\mathbf{\Omega})$, and the winner coalition face competition from other coalition in Stage II, then we have

$$R_{\mathcal{G}} \le R_{win}(\mathbf{\Omega}) \le R'_{win}(\mathbf{\Omega}),$$

where $R'_{win}(\mathbf{\Omega})$ is the profit of the coalition ignoring the competition in Stage II. However, whenever the grand coalition is not formed, coalitions in $\mathbf{\Omega}$ can always form a new coalition with any other coalition to gain more profit. Since when $c_{max} \le 2c_{min}$, any two coalition $\mathcal{B}$ and $\mathcal{B}'$ satisfy

$$c_{\mathcal{B}} \le c_{max} \le 2c_{min} \le 2c_{\mathcal{B}'},$$

which guarantees there is always a coalition would cooperate with the winner coalition in price competition to receive higher profit until the grand coalition forms, i.e.,

$$R_{win'}(\mathbf{\Omega}') \le R_{\mathcal{G}},$$

where $R_{win'}(\mathbf{\Omega}')$ is the profit of winner coalition in any coalition, which is contradictory to our assumptions, and we finish the proof

Consider a coalition structure $\mathbf{\Omega}$ that $\sum_{\mathcal{B}\in\mathbf{\Omega}} \omega_{\mathcal{B}}(\mathbf{\Omega}) < \omega_{\mathcal{G}}$, where $\omega_{\mathcal{G}}$ is the long-term discounted profit of grand coalition. For any coalition structure, the proposer can receive a positive surplus if it proposes a contract to form the grand coalition, since $\omega_{\mathcal{G}} - \sum_{\mathcal{B}\in\mathbf{\Omega}} \hat{\omega}_{\mathcal{B}}(\sigma|\mathbf{\Omega}) > 0$, then the coalition structure would converge to grand coalition in finite periods.

Appendix E
Proof of Theorem 3

According to Theorem 2, CSPs with marginal cost $c_i \le 2c_{min}, \forall i \in \mathcal{S}$ would increase the η for any coalition, then it would be in the stable coalition, since as long as when they are proposers, they can propose a contract to join the only coalition that gains profit.

As long as a CSP can increase the profit of the winner coalition in the price competition would be in the stable coalition. We use E to denote the set of CSPs that form the coalition with the highest η, for any coalition $\mathcal{B}$, we have

$$\eta_{E\backslash\{i\}} \ge \eta_{\mathcal{B}\backslash\{i\}}, \forall i \in E.$$

We need to prove

$$q_i \ge q_{\mathcal{B}\backslash\{i\}}\left(\frac{c_i}{c_{\mathcal{B}\backslash\{i\}}} - 2\right), \tag{2}$$

which can guarantee coalition B could gain higher profit if CSP i join in based on (1). When $q_{\mathcal{B}\backslash\{i\}} < q_{E\backslash\{i\}}, \forall i \in E$, we have

$$q_i \ge q_{E\backslash\{i\}}\left(\frac{c_i}{c_{E\backslash\{i\}}} - 2\right),$$

$$\begin{aligned}
\eta_{E\backslash\{i\}} &> \eta_{\mathcal{B}\backslash\{i\}}, \\
\frac{q_{E\backslash\{i\}}}{q_{\mathcal{B}\backslash\{i\}}} &> \frac{c_{E\backslash\{i\}}}{c_{\mathcal{B}\backslash\{i\}}}, \\
\frac{q_{E\backslash\{i\}}}{q_{\mathcal{B}\backslash\{i\}}} &\ge 1, \\
\frac{c_{E\backslash\{i\}}}{c_{\mathcal{B}\backslash\{i\}}} &\ge 1, \\
c_{E\backslash\{i\}} &\ge c_{\mathcal{B}\backslash\{i\}}.
\end{aligned}$$

Then

$$\frac{q_{E\backslash\{i\}}}{q_{\mathcal{B}\backslash\{i\}}} \ge 1 \ge \frac{\frac{c_i}{c_{E\backslash\{i\}}} - 2}{\frac{c_i}{c_{\mathcal{B}\backslash\{i\}}} - 2}$$

$$\Leftrightarrow q_{E\backslash\{i\}}\left(\frac{c_i}{c_{E\backslash\{i\}}} - 2\right) - q_{\mathcal{B}\backslash\{i\}}\left(\frac{c_i}{c_{\mathcal{B}\backslash\{i\}}} - 2\right) > 0.$$

We have

$$q_i \ge q_{E\backslash\{i\}}\left(\frac{c_i}{c_{E\backslash\{i\}}} - 2\right) \ge q_{\mathcal{B}\backslash\{i\}}\left(\frac{c_i}{c_{\mathcal{B}\backslash\{i\}}} - 2\right) \tag{3}$$

If $q_{\mathcal{B}\backslash\{i\}} \ge q_{E\backslash\{i\}}$,

$$\eta_i \ge \eta_{E\backslash\{i\}} - \frac{2q_{E\backslash\{i\}}}{c_i} \ge \eta_{\mathcal{B}\backslash\{i\}} - \frac{2q_{E\backslash\{i\}}}{c_i} > \eta_{\mathcal{B}\backslash\{i\}} - \frac{2q_{\mathcal{B}\backslash\{i\}}}{c_i},$$

then Eq. (2) is always true, and we finish the proof.

Appendix F
Proof of Theorem 4

If $\delta = 0$, for the winner coalition in the price competition game, its long-term discounted profit is $\hat{\omega}_{win}(\mathbf{\Omega}) = R_{win}(\mathbf{\Omega})$. A newly formed coalition $\mathcal{D}$ satisfies $\eta_{\mathcal{D}} > \eta_{win}$ would be enough since $R_{\mathcal{D}} > R_{win}$,

$$e_{\mathcal{D}}(\mathbf{\Omega}) = \omega_{\mathcal{D}}(\mathbf{\Omega}_{\mathcal{D}}) - \sum_{\mathcal{B}\in\mathcal{D}} \hat{\omega}_{\mathcal{B}}(\mathbf{\Omega}) = R_{\mathcal{D}} - R_{win} > 0,$$

indicating that the contract to form coalition $\mathcal{D}$ can generate a positive surplus.

When $\delta \to 1$, and coalition $\mathcal{D}$ is the stable coalition, we have $\sum_{i\in\mathcal{S}} \hat{\omega}_i(\mathbf{\Omega}) = \hat{\omega}_{\mathcal{D}}(\mathbf{\Omega}_{\mathcal{D}})$. Coalition $\mathcal{D}$ is s the one that gains the most profit, otherwise, it would not be the stable coalition. Since the grand coalition is not the efficient coalition, we have $\omega_{\mathcal{D}}(\mathbf{\Omega}_{\mathcal{D}}) > \omega_{\mathcal{G}}$, and for any coalition structure forming the grand coalition, the surplus is

$$e_{\mathcal{G}}(\mathbf{\Omega}) = \omega_{\mathcal{G}} - \sum_{\mathcal{B}\in\mathbf{\Omega}} \hat{\omega}_{\mathcal{B}}(\mathbf{\Omega}) = \omega_{\mathcal{G}} - \hat{\omega}_{\mathcal{D}}(\mathbf{\Omega}_{\mathcal{D}}) < 0,$$

indicating that the grand coalition would not form.

Considering an efficient coalition $\mathcal{D}$, for any coalition structure $\mathbf{\Omega}$ and $\mathbf{\Omega}_{\mathcal{D}}$, satisfying $\omega_{\mathcal{D}}(\mathbf{\Omega}_{\mathcal{D}}) \ge \sum_{\mathcal{B}\in\mathbf{\Omega}} \omega_{\mathcal{B}}(\mathbf{\Omega})$, then for any $\mathcal{D}_1 \cup \mathcal{D}_2 \cup \cdots \cup \mathcal{D}_n = \mathcal{D}$,

$$\omega_{\mathcal{D}}(\mathbf{\Omega}_{\mathcal{D}}) \ge \sum_{\mathcal{B}\in\mathbf{\Omega}} \omega_{\mathcal{B}}(\mathbf{\Omega}),$$

Algorithm 1: Contract Proposal Algorithm

```
Initialize: networks' weights θ^c and θ^a; target
  networks' weights θ^c' ← θ^c, θ^a' ← θ^a; replay buffer
  E;
for episode = 1 : M do
    Observe the initial coalition structure s_1;
    for t = 1 : T do
        Select contract a_t^* according to (9);
        Observe the current profit R_t;
        Observe the new coalition structure s_{t+1};
        Store experience (s_t, a_t^*, R_t, s_{t+1}) in E;
        Sample a random experience batch N from E;
        Calculate the target value y_m, ∀m ∈ N
          according to (6);
        Update θ^c by minimizing the loss in (7);
        Update θ^a using the sampled gradient in (5);
        Update θ^c' and θ^a' according to (8);
    end
end
```

$$\omega_{\mathcal{D}}(\boldsymbol{\Omega}_{\mathcal{D}}) - (\omega_{D_1} + \cdots \omega_{D_n}) \geq 0,$$

and for any coalition $\mathcal{C} \subset \mathcal{D}$,

$$e_{\mathcal{D}} - e_{\mathcal{C}} = \omega_{\mathcal{D}} - \omega_{\mathcal{C}} - \sum_{\mathcal{B} \in \mathcal{D} \setminus \mathcal{C}} \omega_{\mathcal{B}} > 0, \tag{4}$$

which means for coalitions in $\mathcal{D}$, they would immediately form $\mathcal{D}$, since this contract with the largest surplus.

Appendix G
DDPG Algorithm for CBG

In the following, we first introduce the main idea of the scalable DDPG algorithm. We then introduce the three key parts of the algorithm, i.e., Deep Deterministic Policy Gradient (DDPG), k-nearest-neighbor, and contract evaluation. We finally discuss the convergence of the algorithm.

Main idea of contract proposal algorithm. Inspired by [2], we design a scalable DDPG algorithm for the proposer to calculate the contract that generates the most surplus. The main idea is to use the DDPG algorithm to choose contracts in a continuous space, and use the k-nearest-neighbor methods to derive the corresponding contracts in the discrete space. The training phase of the algorithm consists of M training episodes where each training episode has T time periods. At the beginning of each training episode, we initialize the coalition structure. In each time period, we select a proposer to propose the optimal contract. Specifically, we choose k nearest-neighbor contracts from the actor network's output, then select the best contract by the critic network using the contract evaluation rule. With a well-trained framework, we can calculate the contract for the proposer given any coalition structure.

Deep deterministic policy gradient [3]. DDPG consists of two deep neural networks, i.e., an actor $\boldsymbol{\theta}^a$ and a critic $\boldsymbol{\theta}^c$. In CBG, we use a to represent the contract selected by the actor, and use s to denote the state (i.e., the coalition structure and the proposer coalition). The actor selects action a under state s to maximize the critic's estimated value $Q_{\boldsymbol{\theta}^c}(s,a)$. To enhance stability and efficiency, DDPG incorporates target networks, denoted by $\boldsymbol{\theta}^{a\prime}$ and $\boldsymbol{\theta}^{c\prime}$, to guide the training of the actor network and the critic network, respectively.

At state s, the actor network outputs a proto-contract $\hat{a}$ in the continuous space through function $f_{\boldsymbol{\theta}^a}$ parameterized by $\boldsymbol{\theta}^a$, i.e.,

$$\hat{a} = f_{\boldsymbol{\theta}^a}(s).$$

During the interaction with the environment, the DDPG algorithm accumulates experiences (i.e., state s_t, action a_t^*, reward R_t, next state s_{t+1}) to build an experience replay buffer $\mathcal{E}$, and sample a random experience batch $\mathcal{N}$ from $\mathcal{E}$ to update the actor network's parameters $\boldsymbol{\theta}^a$ through policy gradient methods [4]. This involves approximating the gradient of the function $f_{\boldsymbol{\theta}^a}$ using sampled experiences in set $\mathcal{N}$:

$$\nabla_{\boldsymbol{\theta}^a} f_{\boldsymbol{\theta}^a} \approx \frac{1}{|\mathcal{N}|} \sum_{m \in \mathcal{N}} \nabla_a Q_{\boldsymbol{\theta}^c}(s, \hat{a})|_{\hat{a}=f_{\boldsymbol{\theta}^a}(s_m)} \cdot \nabla_{\boldsymbol{\theta}^a} f_{\boldsymbol{\theta}^a}(s)|_{s_m}. \tag{5}$$

We update the critic network's parameters $\boldsymbol{\theta}^c$ by minimizing the loss between the target value and the critic's estimated value. We denote y_m as the target value in experience $m \in \mathcal{N}$, calculated as:

$$\begin{aligned} &y_m = R_m + \\ &\delta\left(\rho_m \cdot Q_{\boldsymbol{\theta}^{c\prime}}(s_{m+1}, \pi_{\boldsymbol{\theta}^{a\prime}}(s_{m+1})) + (1-\rho_m) Q_{\boldsymbol{\theta}^{c\prime}}(s_m, a_m)\right). \end{aligned} \tag{6}$$

Here ρ_m is the probability that the proposer coalition in experience m is selected as the proposer. We calculate the loss as follows, with which we update $\boldsymbol{\theta}^c$ using policy gradient techniques [3]:

$$L(\boldsymbol{\theta}^c) = \frac{1}{|\mathcal{N}|} \sum_{m \in \mathcal{N}} \left[y_m - Q_{\boldsymbol{\theta}^c}(s_m, a_m)\right]^2. \tag{7}$$

At each time period, we update target networks from local networks, with a controlling parameter λ as follows:

$$\boldsymbol{\theta}^{c\prime} \leftarrow \lambda \boldsymbol{\theta}^c + (1-\lambda)\boldsymbol{\theta}^{c\prime},\ \boldsymbol{\theta}^{a\prime} \leftarrow \lambda \boldsymbol{\theta}^a + (1-\lambda)\boldsymbol{\theta}^{a\prime}. \tag{8}$$

K-nearest-neighbor [5]. The actor network produces a proto-contract $\hat{a}$ in the continuous space, which may be infeasible to the proposer. To solve this problem, we employ the k-nearest-neighbor approach, where we select k nearest-neighbor contracts from the feasible discrete contract set $\mathcal{A}_l$. We denote the set of the k chosen discrete contracts as set $\mathcal{A}_k$, i.e.,

$$\mathcal{A}_k = \arg\overset{k}{\min_{a \in \mathcal{A}_l}} |a - \hat{a}|_2.$$

Contract evaluation. The proposer chooses among the contracts in set $\mathcal{A}_k$ to maximize its surplus, i.e.,

$$a^* = \arg\max_{a \in \mathcal{A}_k} e_{\boldsymbol{\theta}^c}(s, a). \tag{9}$$

Each contract $a \in \mathcal{A}_k$ specifies a set $\mathcal{D}(s,a)$ of coalitions to form a new coalition, and we write $\mathcal{D}(s,a)$ as $\mathcal{D}$ for brevity. The surplus $e_{\boldsymbol{\theta}^c}(s,a)$ is the difference between the estimated long-term discounted profit $Q_{\boldsymbol{\theta}^c}(s_{\mathcal{D}}(\mathcal{D}), a_0(s_{\mathcal{D}}))$ of the new coalition formed by coalitions in set $\mathcal{D}$ and the estimate long-term discounted profits $Q_{\boldsymbol{\theta}^c}(s(\mathcal{B}), a_0(s))$ of coalition $\mathcal{B} \in \mathcal{D}$, calculated as follows:

$$e_{\boldsymbol{\theta}^c}(s, a) = Q_{\boldsymbol{\theta}^c}(s_{\mathcal{D}}(\mathcal{D}), a_0(s_{\mathcal{D}})) - \sum_{\mathcal{B} \in \mathcal{D}(s,a)} Q_{\boldsymbol{\theta}^c}(s(\mathcal{B}), a_0(s)).$$

Here $s_{\mathcal{D}}$ is the coalition structure transformed from s when coalitions in set $\mathcal{D}$ form a new coalition, and $s(\mathcal{B})$ denotes that coalition $\mathcal{B}$ is the proposer in coalition structure s. The term $a_0(s)$ denotes the contract does not form a new coalition in state s (i.e., the contract that does not change the coalition structure).

Convergence analysis. The convergence of DDPG remains an open challenge [6]. The difficulty is mainly due to DDPG's reliance on neural networks, which involve nonlinear function approximation and numerous hyperparameter selections. Furthermore, the training process of DDPG mainly solves a non-convex optimization problem (i.e., minimizing the error of the Q-value function), which typically has multiple local minima, complicating the search for the global minimum. Consequently, there is no convergence guarantee for DDPG. In Section VI, we show the convergence performance of our algorithm through extensive experiments.

Appendix H
Proof of Lemma 3

For CSP $i \in \mathcal{S}$ that $\eta_i < \eta_{max}$, unilaterally joining the intercloud broker does not yield positive profit since $\eta_{\{i\}} \leq \eta_i$, where $\{i\}$ represents the intercloud broker comprising only CSP i. For CSP i with η_{max}, if it joins intercloud broker, and it has $\eta_{\{i\}} > \eta_{sec}$, i.e. CSP i still win in the price competition, however, it has to share its profit with the intercloud broker which leads to profit decreasing. If $\eta_{\{i\}} \leq \eta_{sec}$, CPS i still gain zero profit due to competition.

Appendix I
Proof of Lemma 4

In the intercloud broker model, CSP i's profit is

$$R_i(x_i, \boldsymbol{p}) = \begin{cases} p_i d_i - c_i d_i, & x_i = 0, \\ (p^s(q_i) - c_i)\,\dfrac{q_i}{q_s} d_s, & x_i = 1. \end{cases} \forall i \in \mathcal{S} \quad (10)$$

To maximize the number of CSPs that gain profit, we aim to let CSPs cooperate through the intercloud broker since there is only one SP that could gain positive profit due to the price competition. For CSPs that $p^s(q_i) > c_i, \forall i \in \mathcal{S}$, their cooperation strategies are joining intercloud broker, the rest of CSPs with $\eta \leq \dfrac{1}{K}$ have no incentive to join the intercloud broker, since they would receive negative profit due to the high cost compared to purchase price from the intercloud broker.

Appendix J
Proof of Theorem 5

When

$$R_i^{coop} \sum_{j=1}^{\infty} \delta_i^j \geq R_i^{coop} \sum_{j=1}^{t-1} \delta_i^j + R_i^{dev} \delta_i^{t-1} + R_i^{pun} \sum_{j=t+1}^{\infty} \delta_i^{j-1},$$

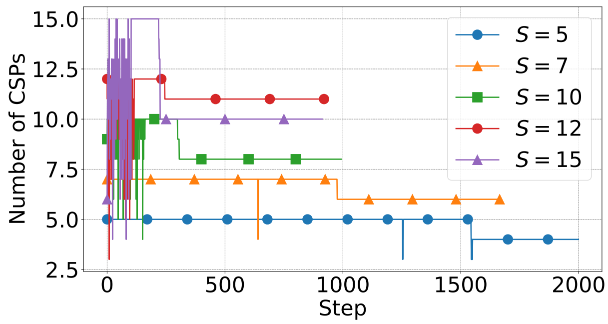

Fig. 1: Number of cooperative CSP under different number of CSP S in market

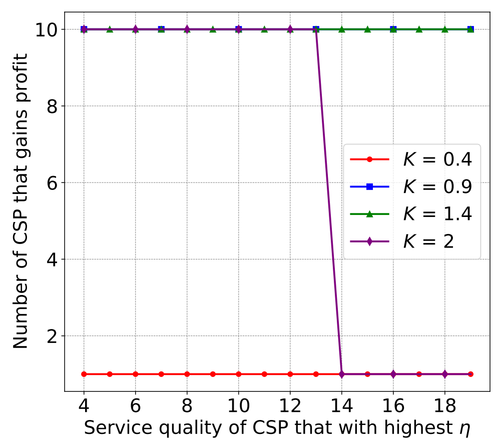

Fig. 2: Number of CSP that receives positive profit under different service quality of CSP with the highest η

CSP i would not deviate from the cooperation at time slot t. And

$$R_i^{coop} \sum_{j=1}^{t-1} \delta_i^j + R_i^{dev} \delta_i^{t-1} + R_i^{pun} \sum_{j=t+1}^{\infty} \delta_i^{j-1}$$
$$= R_i^{coop} \frac{1 - \delta_i^{t-1}}{1 - \delta_i} + R_i^{dev} \delta_i^{t-1} + R_i^{pun} \frac{\delta_i^t}{1 - \delta_i},$$

$$R_i^{coop} \sum_{j=1}^{\infty} \delta_i^j = R_i^{coop} \frac{1}{1 - \delta_i},$$

then for CSPs $\forall i \in \mathcal{S}$ with $\delta_i \geq \dfrac{R_i^{dev} - R_i^{coop}}{R_i^{dev} - R_i^{pun}}$, they would play cooperation strategies $\boldsymbol{x}^{coop}$ at each stage, and would not deviate from cooperation in the long run.

Appendix K
Additional Simulation Analysis

We first use an example to show the stable coalition of CSPs' dynamic coalition formation process. In the example, CSPs' service qualities and costs are as follows

$$\boldsymbol{q} = [2.96, 2.60, 1.92, 2.56, 1.24, 2.28, 1.29, 2.89, 2.04, 1.83],$$

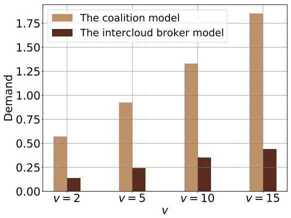

Fig. 3: Users' demand under two cooperation models

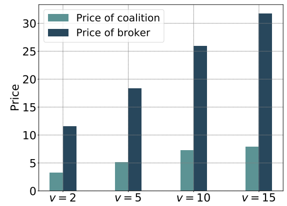

Fig. 4: Price under two cooperation models

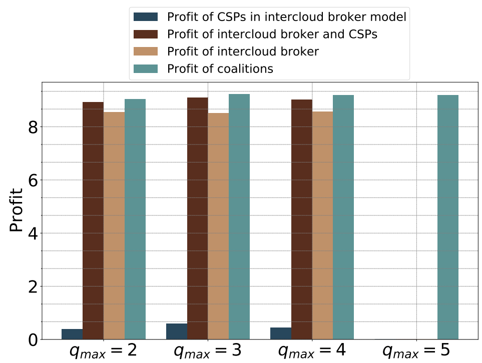

Fig. 5: Profits in two cooperation models under different q_{max}

$$\boldsymbol{c} = [0.24, 0.20, 0.69, 0.22, 0.28, 0.43, 0.84, 0.19, 0.85, 0.19].$$

In this case, CSP 7 and CSP 9 are stand-alone CSPs, while other CSPs choose to cooperate and form a coalition, which is the stable coalition. We can see that CSP 7 and CSP 9 who are not in the coalition have high marginal costs. This implies that CSPs with low marginal costs form the stable coalition, which verifies our theoretical result in Theorem 3.

We show that in different numbers of CSPs in the coalition model in Fig. 1. There is a regularity that the grand coalition would form in the early training steps. This is because we set the k of k-nearest-neighbor as 10, and we stop choosing a proposer to form a new coalition only when the grand coalition is formed, or we have chosen a proposer for over 50 times. When there are 5 CSPs, the proposer has an action space of 2^5 combinations, and the k-nearest-neighbor would be more likely to form a coalition with more CSPs since the larger coalition tends to be more efficient than smaller ones.

We show the number of CSPs that gain profit in Fig. 2, as the service quality of CSP with the maximum η increases. When there is a large difference among CSPs, i.e. the service quality of CSP with the maximum η is high, the remaining CSPs would not join the broker, and the broker can not attract enough CSPs to generate positive profit when intercloud broker sets a low price (K is small) for CSPs. However, when $K = 2$, as the service quality of CSP with the maximum η increases, the intercloud broker still can not attract CSPs, since CSP with the maximum η can gain more profit to be stand-alone SP even though other CSPs all join in intercloud broker than all CSPs join in the intercloud broker.

We compare users' demands and the service prices of the two cooperation models in Fig. 3 and Fig. 4, respectively. We can see that compared to the intercloud broker model, forming a coalition between CSPs directly leads to a lower service price, and thus users have more computing demands. Compared to the coalition model, the intercloud broker sets a higher service price to offset the costs due to purchasing computing services from CSPs, which results in fewer computing demands from users.

We compare the profits with different maximum service qualities q_{max} of CSPs under the two cooperation models in Fig. 5, where $K = 0.5$. We can see that compared to the intercloud broker model, the coalition obtains a higher profit. The reason is that the intercloud broker needs to purchase computing services from CSPs, and hence its quality-cost ratio is small, which results in low computing demands from users. Furthermore, when q_{max} is large, the intercloud broker cannot attract any CSP and obtains zero profit.

REFERENCES

[1] A. Gomes, "Multilateral contracting with externalities," *Econometrica*, vol. 73, no. 4, pp. 1329–1350, 2005.
[2] G. Dulac-Arnold, R. Evans, H. van Hasselt, P. Sunehag, T. Lillicrap, J. Hunt, T. Mann, T. Weber, T. Degris, and B. Coppin, "Deep reinforcement learning in large discrete action spaces," *arXiv preprint arXiv:1512.07679*, 2015.
[3] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2016*, 2016.
[4] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *International conference on machine learning*. PMLR, 2014, pp. 387–395.
[5] N. Papernot and P. Mcdaniel, "Deep k-nearest neighbors: Towards confident, interpretable and robust deep learning," *arXiv preprint arXiv:1803.04765*, 2018.
[6] H. Xiong, T. Xu, L. Zhao, Y. Liang, and W. Zhang, "Deterministic policy gradient: Convergence analysis," in *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, vol. 180. PMLR, 01–05 Aug 2022, pp. 2159–2169.