APPENDIX

We primarily rely on the *λ-Lipschitz* and *Brouwer*'s fixed point theorem to prove the existence of *NE* for the P2P energy trading based on stochastic game.

In the stochastic game process of P2P energy trading, considering the interaction of actions among prosumers, consistent with MDP, Eq. (11) can be reformulated into the following two forms via Bellman's equivalence equation:

$$\begin{aligned} & V_{\psi_n}(s) \\ & = \sum_{a_n^t \in \boldsymbol{A}_n} \psi_n(a_n^t \mid s)(R_s^{(a_n^t, \Psi_{-n})} + \gamma \sum_{s^{'} \in \boldsymbol{S}_n} P^{(s^{'}, a_n^t)}(s^{'} \mid s, a_n^t) V_{\psi_n}(s^{'})) \end{aligned} \tag{A1}$$

$$Q_{\psi_n}(a_n^t, \boldsymbol{\Psi}_{-n}) = R_s^{(a_n^t, \Psi_{-n})} + \gamma(\sum_{s^{'} \in \boldsymbol{S}_n} P^{(s^{'}, a_n^t)}(s^{'} \mid s, a_n^t) V_{\psi_n}(s^{'})) \tag{A2}$$

where $V_{\psi_n}(s)$ 、 $Q_{\psi_n}(a_n^t, \boldsymbol{\Psi}_{-n})$ denote the state-value function and action-value function of Eq.(11), respectively; $R_s^{(a_n^t, \Psi_{-n})}$ represents the reward in the current state of prosumer *n*; $V_{\psi_n}(s^{'})$ indicates the reward in state $s^{'}$ of prosumer *n*; γ is the discount factor, in this paper $\gamma = 1$.

Then we define a function as follows:

$$\varphi_n(\psi) = \max\left\{0, Q_{\psi_n}(\tilde{a}_n^t, \boldsymbol{\Psi}_{-n}) - V_{\Psi_n}(s)\right\} \tag{A3}$$

Eq.(A3) illustrates the influence of prosumer *n* on self-reward only when prosumer *n* changes individual action. If there is invariably $\varphi_n(\psi) = 0$ for all prosumers, it indicates that the stochastic game is in *NE*, and Ψ represents an optimal strategy.

Furthermore, an auxiliary function *f* is constructed based on $\varphi_n(\psi)$:

$$f_n(\psi) = \frac{\psi_n(s_n^t, \tilde{a}_n^t) + \varphi_n(\psi)}{1 + \sum_{b_n^t \in \boldsymbol{A}_n} \varphi_n(\psi)} \tag{A4}$$

The auxiliary function maps the original strategy ψ_n to a new strategy $f_n(\psi)$. If an action of the prosumer *n* in the new strategy can improve the reward, the probability of that corresponding action also increases.

we first proof the *λ-Lipschitz* continuity of the $f_n(\psi)$. In the P2P trading process, as the strategies of each prosumer interact with one another, the set $\boldsymbol{\Psi}$, composed of all mixed strategies, can be expressed as the Cartesian product of strategies of all prosumers:

$$\boldsymbol{\Psi} = \boldsymbol{\Psi}_1 \times \boldsymbol{\Psi}_2 \times \cdots \times \boldsymbol{\Psi}_n \times \cdots \times \boldsymbol{\Psi}_N \tag{A5}$$

Similarly, the set of actions for all prosumers can be expressed as:

$$\boldsymbol{A} = \boldsymbol{A}_1 \times \boldsymbol{A}_2 \times \cdots \times \boldsymbol{A}_n \times \cdots \times \boldsymbol{A}_N \tag{A6}$$

Since all the P2P trading strategies of each prosumer are finite sets, the infinite norm distance for two arbitrary strategies in $\boldsymbol{\Psi}$ can be expressed as:

$$\left\|\boldsymbol{\Psi}^1 - \boldsymbol{\Psi}^2\right\|_\infty \le \max\left|\boldsymbol{\Psi}^1(s_n, a_n) - \boldsymbol{\Psi}^2(s_n, a_n)\right| = \delta \tag{A7}$$

where δ is the maximum infinite norm distance.

It is sufficient to prove the continuity property of $f_n(\psi)$ by demonstrating that arbitrarily strategies

$\boldsymbol{\Psi}^1$, $\boldsymbol{\Psi}^2$ can be used to make $\left\| f(\boldsymbol{\Psi}^1) - f(\boldsymbol{\Psi}^2) \right\|_\infty \le \lambda\delta$ that satisfies the λ-*Lipschitz* continuity condition.

For this purpose, we draw the theoretical support from the following Lemma in [37].

Lemma 1: The function $f_n(\psi)$ is λ-*Lipschitz*, since for any $\boldsymbol{\Psi}^1, \boldsymbol{\Psi}^2 \in \boldsymbol{\Psi}$,we have

$$\left\| f(\boldsymbol{\Psi}^1) - f(\boldsymbol{\Psi}^2) \right\|_\infty \le \frac{11NS^2 A_{\max}^2 r_{\max}}{(1-\gamma)^2}\delta \tag{A8}$$

where S is the state dimensions of prosumer; $A_{\max}$ is the maximum dimensions of the action space for prosumers; $r_{\max}$ is the maximization of profits for prosumers in all time periods.

It is clear that $f_n(\psi)$ exhibits λ-*Lipschitz* continuity. Furthermore, $\psi_n \in [0,1]$ and $f_n(\psi) \in [0,1]$ contribute to $f_n(\psi)$ being a function of $[0,1] \to [0,1]$.Thus, the existence of *NE* in the stochastic game can be proved through *Brouwer*'s fixed point theorem. We just require to demonstrate that a P2P trading strategy ψ_n is the *NE* of the stochastic game if and only if ψ_n is a fixed point of $f_n(\psi)$.

1) The Proof of Necessity

We assume that ψ_n^* is an *NE* of the stochastic game. It is obtained by applying ψ_n^* to Eq.(A3):

$$\varphi_n(\psi_n^*) = \max\left\{0, Q_{\psi_n}(\tilde{a}_n^t, \boldsymbol{\Psi}_{-n}^*) - Q_{\Psi_n}(\psi_n^*, \boldsymbol{\Psi}_{-n}^*)\right\} \tag{A9}$$

From Eq.(A9), we can note that $\varphi_n(\psi_n^*) \equiv 0$. This is because that ψ_n^* is the *NE* of the stochastic game, and regardless of any changes in actions made by prosumer *n*, the $Q_{\psi_n}(\tilde{a}_n^t, \boldsymbol{\Psi}_{-n}^*) \le Q_{\Psi_n}(\psi_n^*, \boldsymbol{\Psi}_{-n}^*)$ .Therefore, Eq.(A4) can be expressed as follows:

$$f_n(\psi_n^*) = \frac{\psi_n^*(s_n^t, \tilde{a}_n^t) + \varphi_n(\psi_n^*)}{1 + \sum_{b_n^t \in \boldsymbol{A}_n} \varphi_n(\psi_n^*)} = \psi_n^* \tag{A10}$$

Hence, the *NE* of the stochastic game is the fixed point of $f_n(\psi)$.

2) The Proof of Sufficiency

We suppose that ψ_n is the fixed point of $f_n(\psi)$, thus:

$$f_n(\psi_n) = \frac{\psi_n(s_n^t, a_n^t) + \varphi_n(\psi_n)}{1 + \sum_{b_n^t \in \boldsymbol{A}_n} \varphi_n(\psi_n)} = \psi_n \tag{A11}$$

At the fixed point ψ_n , $Q_{\psi_n}(a_n^t, \boldsymbol{\Psi}_{-n}) \le V_{\psi_n}(s)$ holds for any action of the prosumer *n*. To verify this assertion, let's assume the existence of $c_n \in \boldsymbol{A}$ such that $Q_{\psi_n}(c_n, \boldsymbol{\Psi}_{-n}) > V_{\psi_n}(s)$. Consequently, we derive $\psi_n(s, c_i) > 0$.

Afterwards, Eq.(A1) is further derived using the Bellman policy equation:

$$
\begin{aligned}
&V_{\psi_n}(s) \\
&= \sum_{a_n^t \in \boldsymbol{A}_n} \psi_n(s, a_n^t)(R_s^{(a_n^t, \Psi_{-n})} + \gamma \sum_{s' \in \boldsymbol{S}_n} P^{(s', a_n^t)}(s' \mid s, a_n^t) V_{\psi_n}(s')) \\
&= \sum_{a_n^{t'} \in \boldsymbol{A}_n^+} \psi_n(s, a_n^{t'})(Q_{\psi_n}(a_n^{t'}, \boldsymbol{\Psi}_{-n}))
\end{aligned} \tag{A12}
$$

where $\boldsymbol{A}_n^+$ is a set of action in which $\psi_n(s, a_n^{t'}) > 0$.

As $\sum_{a_n^{t'} \in \boldsymbol{A}_n^+} \psi_n(s, a_n^{t'}) = 1$, there must exists action $d_n \in \boldsymbol{A}_n^+$ such that $(Q_{\psi_n}(a_n^{t'}, \boldsymbol{\Psi}_{-n})) < V_{\psi_n}(s)$ .It can be obtained by applying d_n to Eq.(A11) :

$$
\begin{aligned}
f_n(\psi_n(s, d_n)) &= \frac{\psi_n(s_n^t, d_n) + \varphi_n(\psi_n)}{1 + \sum_{b_n^t \in \boldsymbol{A}_n} \varphi_n(\psi_n)} \\
&= \frac{\psi_n(s_n^t, d_n)}{1 + \sum_{b_n^t \in \boldsymbol{A}_n} \varphi_n(\psi_n)} \\
&\le \frac{\psi_n(s_n^t, d_n)}{1 + Q_{\psi_n}(c_n, \boldsymbol{\Psi}_{-n}) - V_{\psi_n}(s)} \neq \psi_n(s, d_n)
\end{aligned} \tag{A13}
$$

From Eq.(A13), it follows that if it is assumed that there is $Q_{\psi_n}(c_n, \boldsymbol{\Psi}_{-n}) > V_{\psi_n}(s)$ at the fixed point ψ_n, the result obtained contradicts the *Brouwer*'s fixed point theorem. Therefore, $Q_{\psi_n}(a_n^t, \boldsymbol{\Psi}_{-n}) \le V_{\psi_n}(s)$ exists at the fixed point.

Hence Eq.(A11) can be expressed as

$$
\psi_n = \frac{\psi_n(s_n^t, a_n^t)}{1 + \sum_{b_n^t \in \boldsymbol{A}_n} \varphi_n(\psi_n)} \tag{A14}
$$

In the aforementioned equation, it is evident that the numerator remains consistent on both sides of the equation. Consequently, the denominator must equate to 1. This signifies that for prosumer *n* and $b_n^t \in \boldsymbol{A}_n$, $\sum_{b_n^t \in \boldsymbol{A}_n} \varphi_n(\psi_n) = 0$. This observation suggests that no prosumer can improve its profits by altering its strategy, thus indicating the optimality of the strategy in this context, which corresponds to the *NE* of the stochastic game.

Therefore, the fixed point ψ_n serves as the *NE* (ψ_n^*) of the stochastic game.

So far, the proof of the existence of *NE* in the stochastic game is complete.