PROOF OF THE EQUIVALENCE BETWEEN THE ORIGINAL PROBLEM AND DECOMPOSED PROBLEMS

In order to prove the equivalence between the original problem and decomposed problems, the following two theorems are proved at first.

Theorem 1: Let $\boldsymbol{a}_{k,t}^{fu}$ be the optimal solution of subproblem $\boldsymbol{SP}_{k,t}^{F}$, given system state $\boldsymbol{S}_t$ and leader's decision $\boldsymbol{a}_t^l$ at period *t*. Then, the decision $\left(\boldsymbol{a}_{k,t}^{fu},\boldsymbol{a}_{k,t+1}^{f*},\cdots,\boldsymbol{a}_{k,T-1}^{f*},\boldsymbol{a}_{k,T}^{f*}\right)$ will be the best response to the leader's decision $\left(\boldsymbol{a}_t^l,\boldsymbol{a}_{t+1}^{l*},\cdots,\boldsymbol{a}_{T-1}^{l*},\boldsymbol{a}_T^{l*}\right)$ with the system state $\boldsymbol{S}_t$ at period *t*. Mathematically, it can be described as (1) and (2).

$$\boldsymbol{R}_{k,t}^{f}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l\right)=\underset{\boldsymbol{a}_{k,t}^{f}\in\boldsymbol{\Pi}_{k,t}^{f}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l\right)}{\arg\min}\left\{g_{k,t}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{a}_{k,t}^{f}\right)+\underset{\xi_{t+1}}{\mathbb{E}}\left[V_{k,t+1}^{f}\left(\boldsymbol{S}_{t+1}\right)\right]\right\}\quad\forall k\in\mathcal{K},\forall t\in\mathcal{T}\quad(1)$$

$$T_{k,t}^{f}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l\right)=\min_{\boldsymbol{a}_{k,t}^{f}\in\boldsymbol{\Pi}_{k,t}^{f}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l\right)}\left\{g_{k,t}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{a}_{k,t}^{f}\right)+\mathbb{E}\left[V_{k,t+1}^{f}\left(\boldsymbol{S}_{t+1}\right)\right]\right\}\quad\forall k\in\mathcal{K},\forall t\in\mathcal{T}\quad(2)$$

where $\boldsymbol{R}_{k,t}^{f}$ denotes the follower *k*'s response function at period *t*; $T_{k,t}^{f}$ denotes the follower *k*'s ancillary function at period *t*; The left side of (1) represents the optimal decision of follower *k* at period *t*, determined by the original problem according to *Definition 1*; The right side of (1) represents the optimal solution of subproblem $\boldsymbol{SP}_{k,t}^{F}$ at period *t*.

Proof of Theorem 1: The theorem is proved by contradiction. Suppose that follower *k*'s decision $\left(\boldsymbol{a}_{k,t}^{fu},\boldsymbol{a}_{k,t+1}^{f*},\cdots,\boldsymbol{a}_{k,T-1}^{f*},\boldsymbol{a}_{k,T}^{f*}\right)$ is not the best response to the leader's decision $\left(\boldsymbol{a}_t^l,\boldsymbol{a}_{t+1}^{l*},\cdots,\boldsymbol{a}_{T-1}^{l*},\boldsymbol{a}_T^{l*}\right)$. Then, there must exist another decision $\left(\boldsymbol{a}_{k,t}^{f\Delta},\boldsymbol{a}_{k,t+1}^{f\Delta},\cdots,\boldsymbol{a}_{k,T-1}^{f\Delta},\boldsymbol{a}_{k,T}^{f\Delta}\right)$, which is the best response to the leader's decision $\left(\boldsymbol{a}_t^l,\boldsymbol{a}_{t+1}^{l*},\cdots,\boldsymbol{a}_{T-1}^{l*},\boldsymbol{a}_T^{l*}\right)$. That is

$$\begin{aligned}&g_{k,t}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{a}_t^{f\Delta}\right)+\underset{\xi_{t+1}\cdots\xi_T}{\mathbb{E}}\left[\sum_{\tau=t+1}^{T}g_{k,\tau}\left(\boldsymbol{S}_\tau^{\Delta},\boldsymbol{a}_\tau^{l*},\boldsymbol{a}_\tau^{f\Delta}\right)\right]\\&\le g_{k,t}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{a}_t^{fu}\right)+\underset{\xi_{t+1}\cdots\xi_T}{\mathbb{E}}\left[\sum_{\tau=t+1}^{T}g_{k,\tau}\left(\boldsymbol{S}_\tau^{u},\boldsymbol{a}_\tau^{l*},\boldsymbol{a}_\tau^{f*}\right)\right]\end{aligned}\quad(3)$$

where the follower *k*'s decision $\boldsymbol{a}_{k,\tau}^{f\Delta}$ requires to satisfy constraints $\boldsymbol{a}_{k,\tau}^{f\Delta}\left(\boldsymbol{S}_\tau^{\Delta},\boldsymbol{a}_t^{l*}\right)\in\boldsymbol{\Pi}_{k,\tau}^{f}\left(\boldsymbol{S}_\tau^{\Delta},\boldsymbol{a}_t^{l*}\right),\tau=t,t+1,\cdots,T$ and $\boldsymbol{S}_t^{\Delta}=\boldsymbol{S}_t$.

Replace the same part in $\left(\boldsymbol{a}_{k,t}^{f\Delta},\boldsymbol{a}_{k,t+1}^{f\Delta},\cdots,\boldsymbol{a}_{k,T-1}^{f\Delta},\boldsymbol{a}_{k,T}^{f\Delta}\right)$ with $\left(\boldsymbol{a}_{k,t+1}^{f*}\left(\boldsymbol{S}_{t+1}^{\Delta1},\boldsymbol{a}_{t+1}^{l*}\right),\boldsymbol{a}_{k,t+2}^{f*}\left(\boldsymbol{S}_{t+2}^{\Delta1},\boldsymbol{a}_{t+2}^{l*}\right),\cdots,\boldsymbol{a}_{k,T}^{f*}\left(\boldsymbol{S}_{T}^{\Delta1},\boldsymbol{a}_{T}^{l*}\right)\right)$, in which $\boldsymbol{S}_{t+1}^{\Delta1}$ is equal to $\boldsymbol{S}_{t+1}^{\Delta}$ and $\boldsymbol{a}_{k,\tau}^{f*}\left(\boldsymbol{S}_\tau^{\Delta1},\boldsymbol{a}_\tau^{l*}\right)$ represents the follower *k*'s optimal decision, given system state $\boldsymbol{S}_\tau^{\Delta1}$ and leader's decision $\boldsymbol{a}_\tau^{l*}$ at period τ. Then, a new decision $\left(\boldsymbol{a}_{k,t}^{f\Delta},\boldsymbol{a}_{k,t+1}^{f*},\cdots,\boldsymbol{a}_{k,T-1}^{f*},\boldsymbol{a}_{k,T}^{f*}\right)$ can be obtained, which satisfies

$$\begin{aligned}&g_{k,t}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{a}_t^{f\Delta}\right)+\underset{\xi_{t+1}\cdots\xi_T}{\mathbb{E}}\left[\sum_{\tau=t+1}^{T}g_{k,\tau}\left(\boldsymbol{S}_\tau^{\Delta},\boldsymbol{a}_\tau^{l*},\boldsymbol{a}_\tau^{f\Delta}\right)\right]\\&\ge g_{k,t}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{a}_t^{f\Delta}\right)+\underset{\xi_{t+1}\cdots\xi_T}{\mathbb{E}}\left[\sum_{\tau=t+1}^{T}g_{k,\tau}\left(\boldsymbol{S}_\tau^{\Delta1},\boldsymbol{a}_\tau^{l*},\boldsymbol{a}_\tau^{f*}\left(\boldsymbol{S}_\tau^{\Delta1},\boldsymbol{a}_\tau^{l*}\right)\right)\right]\\&=g_{k,t}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{a}_t^{f\Delta}\right)+\underset{\xi_{t+1}}{\mathbb{E}}\left[V_{k,t+1}^{f}\left(\boldsymbol{S}_{t+1}^{\Delta}\right)\right]\\&\ge\min_{\boldsymbol{a}_{k,t}^{f}\in\boldsymbol{\Pi}_{k,t}^{f}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l\right)}\left\{g_{k,t}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{a}_t^{f}\right)+\underset{\xi_{t+1}}{\mathbb{E}}\left[V_{k,t+1}^{f}\left(\boldsymbol{S}_{t+1}\right)\right]\right\}\\&=g_{k,t}\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{a}_t^{fu}\right)+\underset{\xi_{t+1}\cdots\xi_T}{\mathbb{E}}\left[\sum_{\tau=t+1}^{T}g_{k,\tau}\left(\boldsymbol{S}_\tau^{u},\boldsymbol{a}_\tau^{l*},\boldsymbol{a}_\tau^{f*}\right)\right]\end{aligned}\quad(4)$$

where the first inequality is derived from *Definition 1*; The second inequality is derived from the characteristic of the min operator.

This leads to a contradiction that there exists a feasible decision $\left(\boldsymbol{a}_{k,t}^{f\Delta},\boldsymbol{a}_{k,t+1}^{f\Delta},\cdots,\boldsymbol{a}_{k,T-1}^{f\Delta},\boldsymbol{a}_{k,T}^{f\Delta}\right)$, which is better than the decision $\left(\boldsymbol{a}_{k,t}^{fu},\boldsymbol{a}_{k,t+1}^{f*},\cdots,\boldsymbol{a}_{k,T-1}^{f*},\boldsymbol{a}_{k,T}^{f*}\right)$. Therefore, by reductio, *Theorem 1* is proved to be true.

End of *Proof 1*.

Theorem 2: Let $\left(\boldsymbol{a}_t^{lu},\boldsymbol{u}_t^{lu}\right)$ be the optimal solution of subproblem $\boldsymbol{SP}_t^{L}$, given system state $\boldsymbol{S}_t$ at period *t*. Then, the decision $\left(\boldsymbol{a}_t^{lu},\boldsymbol{u}_t^{lu},\boldsymbol{a}_{t+1}^{l*},\boldsymbol{u}_{t+1}^{l*},\cdots,\boldsymbol{a}_T^{l*},\boldsymbol{u}_T^{l*}\right)$ will be the optimal decisions of the leader and followers with the system state $\boldsymbol{S}_t$ at period *t*. Mathematically, it can be described as (5).

$$V_t^l\left(\boldsymbol{S}_t\right)=\min_{\substack{\boldsymbol{a}_t^l\in\boldsymbol{\Pi}_t^l\left(\boldsymbol{S}_t\right)\\ \boldsymbol{u}_t^f(k)=\boldsymbol{R}_k\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l\right),\forall k\in\mathcal{K}}}\left\{G_t\left(\boldsymbol{S}_t,\boldsymbol{a}_t^l,\boldsymbol{u}_t^f\right)+\underset{\xi_{t+1}}{\mathbb{E}}\left[V_{t+1}^{l}\left(\boldsymbol{S}_{t+1}\right)\right]\right\}\quad(5)$$

where V_t^l denotes the leader's value function at period *t*; The left side of (5) represents the optimal cost of the leader from period *t* to *T*, determined by the original problem according to *Definition 2*; The right side of (5) represents the optimal solution of subproblem $\boldsymbol{SP}_t^{L}$ at period *t*.

Proof of Theorem 2: The theorem is also proved by contradiction. Suppose that $\left(\boldsymbol{a}_t^{lu},\boldsymbol{u}_t^{lu},\boldsymbol{a}_{t+1}^{l*},\boldsymbol{u}_{t+1}^{l*},\cdots,\boldsymbol{a}_T^{l*},\boldsymbol{u}_T^{l*}\right)$ is not the optimal decisions of the leader and followers with the system state $\boldsymbol{S}_t$ at period *t*. Then, there would exist another decision $\left(\boldsymbol{a}_t^{l\theta},\boldsymbol{u}_t^{f\theta},\boldsymbol{a}_{t+1}^{l\theta},\boldsymbol{u}_{t+1}^{f\theta}\cdots,\boldsymbol{a}_T^{l\theta},\boldsymbol{u}_T^{f\theta}\right)$, which satisfies

$$G_t\left(\boldsymbol{S}_t,\boldsymbol{a}_t^{l\theta},\boldsymbol{u}_t^{f\theta}\right)+\underset{\xi_{t+1}\cdots\xi_T}{\mathbb{E}}\left[\sum_{\tau=t+1}^{T}G_\tau\left(\boldsymbol{S}_\tau^{\theta},\boldsymbol{a}_\tau^{l\theta},\boldsymbol{u}_t^{f\theta}\right)\right]$$

$$\leq G_t\left(\boldsymbol{S}_t, \boldsymbol{a}_t^{lu}, \boldsymbol{a}_t^{fu}\right) + \mathop{\mathbb{E}}_{\xi_{t+1}\cdots\xi_T}\left[\sum_{\tau=t+1}^{T} G_\tau\left(\boldsymbol{S}_\tau^{u}, \boldsymbol{a}_\tau^{l*}, \boldsymbol{a}_\tau^{f*}\right)\right] \tag{6}$$

where the leader's decision requires satisfying constraints $\boldsymbol{a}_\tau^{l\theta}\left(\boldsymbol{S}_\tau^{\theta}\right) \in \boldsymbol{\Pi}_\tau^{l}\left(\boldsymbol{S}_\tau^{\theta}\right), \tau = t, t+1, \cdots, T$, and $\boldsymbol{S}_t^{\theta} = \boldsymbol{S}_t$.

Similar to *Proof 1*, replace the same part in $\left(\boldsymbol{S}_t, \boldsymbol{a}_t^{l\theta}, \boldsymbol{u}_t^{f\theta}, \boldsymbol{S}_{t+1}^{\theta}, \boldsymbol{a}_{t+1}^{l\theta}, \boldsymbol{u}_{t+1}^{f\theta} \cdots, \boldsymbol{S}_T^{\theta}, \boldsymbol{a}_T^{l\theta}, \boldsymbol{u}_T^{f\theta}\right)$ with $\left(\boldsymbol{a}_{t+1}^{l*}\left(\boldsymbol{S}_{t+1}^{\theta 1}\right),\right.$ $\left.\boldsymbol{u}_{t+1}^{f*}\left(\boldsymbol{S}_{t+1}^{\theta 1}, \boldsymbol{a}_{t+1}^{l*}\right), \cdots, \boldsymbol{a}_T^{l*}\left(\boldsymbol{S}_T^{\theta 1}\right), \boldsymbol{u}_T^{f*}\left(\boldsymbol{S}_T^{\theta 1}, \boldsymbol{a}_T^{l*}\right)\right)$, in which $\boldsymbol{S}_{t+1}^{\theta 1}$ is equal to $\boldsymbol{S}_{t+1}^{\theta}$ and $\boldsymbol{a}_\tau^{l*}\left(\boldsymbol{S}_\tau^{\theta 1}\right)$ represents the leader's optimal decision, given system state $\boldsymbol{S}_\tau^{\theta 1}$ at period τ . Then, a new decision $\left(\boldsymbol{S}_t, \boldsymbol{a}_t^{l\theta}, \boldsymbol{u}_t^{f\theta}, \boldsymbol{S}_{t+1}^{\theta 1}, \boldsymbol{a}_{t+1}^{l*}, \boldsymbol{u}_{t+1}^{f*} \cdots, \boldsymbol{S}_T^{\theta 1}, \boldsymbol{a}_T^{l*}, \boldsymbol{u}_T^{f*}\right)$ can be obtained, which satisfies

$$\begin{aligned}
& G_t\left(\boldsymbol{S}_t, \boldsymbol{a}_t^{l\theta}, \boldsymbol{u}_t^{f\theta}\right) + \mathop{\mathbb{E}}_{\xi_{t+1}\cdots\xi_T}\left[\sum_{\tau=t+1}^{T} G_\tau\left(\boldsymbol{S}_\tau^{\theta}, \boldsymbol{a}_\tau^{l\theta}, \boldsymbol{u}_t^{f\theta}\right)\right] \\
& \geq G_t\left(\boldsymbol{S}_t, \boldsymbol{a}_t^{l\theta}, \boldsymbol{u}_t^{f\theta}\right) + \mathop{\mathbb{E}}_{\xi_{t+1}\cdots\xi_T}\left[\sum_{\tau=t+1}^{T} G_\tau\left(\boldsymbol{S}_\tau^{\theta 1}, \boldsymbol{a}_\tau^{l*}\left(\boldsymbol{S}_\tau^{\theta 1}\right), \boldsymbol{u}_\tau^{f*}\left(\boldsymbol{S}_\tau^{\theta 1}, \boldsymbol{a}_\tau^{l*}\right)\right)\right] \\
& = G_t\left(\boldsymbol{S}_t, \boldsymbol{a}_t^{l\theta}, \boldsymbol{u}_t^{f\theta}\right) + \mathop{\mathbb{E}}_{\xi_{t+1}}\left[V_{t+1}^{l}\left(\boldsymbol{S}_{t+1}^{\theta}\right)\right] \\
& \geq \min_{\substack{\boldsymbol{a}_t^{l} \in \boldsymbol{\Pi}_t^{l}(\boldsymbol{S}_t) \\ \boldsymbol{u}_t^{f}(k) = \boldsymbol{R}_k\left(\boldsymbol{S}_t, \boldsymbol{a}_t^{l}\right), \forall k \in \mathcal{K}}} \left\{G_t\left(\boldsymbol{S}_t, \boldsymbol{a}_t^{l}, \boldsymbol{u}_t^{f}\right) + \mathop{\mathbb{E}}_{\xi_{t+1}}\left[V_{t+1}^{l}\left(\boldsymbol{S}_{t+1}\right)\right]\right\} \\
& = G_t\left(\boldsymbol{S}_t, \boldsymbol{a}_t^{lu}, \boldsymbol{a}_t^{fu}\right) + \mathop{\mathbb{E}}_{\xi_{t+1}\cdots\xi_T}\left[\sum_{\tau=t+1}^{T} G_\tau\left(\boldsymbol{S}_\tau^{u}, \boldsymbol{a}_\tau^{l*}, \boldsymbol{a}_\tau^{f*}\right)\right]
\end{aligned} \tag{7}$$

This leads to a contradiction that there exists a feasible decision $\left(\boldsymbol{a}_t^{l\theta}, \boldsymbol{u}_t^{f\theta}, \boldsymbol{a}_{t+1}^{l\theta}, \boldsymbol{u}_{t+1}^{f\theta} \cdots, \boldsymbol{a}_T^{l\theta}, \boldsymbol{u}_T^{f\theta}\right)$, which is better than the decision $\left(\boldsymbol{a}_t^{lu}, \boldsymbol{u}_t^{lu}, \boldsymbol{a}_{t+1}^{l*}, \boldsymbol{u}_{t+1}^{l*}, \cdots, \boldsymbol{a}_T^{l*}, \boldsymbol{u}_T^{l*}\right)$. Therefore, by reductio, *Theorem 2* is proved to be true.

End of *Proof 2*.

According to *Theorem 1* and *Theorem 2*, the optimal solution of subproblems (i.e., $\boldsymbol{SP}_{k,t}^{F}$ and $\boldsymbol{SP}_t^{L}$) is also optimal for the original problem. This indicates that the dynamic Stackelberg problem before and after decomposition are equivalent in mathematics.