I. The Proof of Throughput Optimality of QueueFlower

In this section, we prove the weak stability of QueueFlower. There exists a well-tuned w such that the virtual queue is quite close to the real queue. Based on this, we show the system achieves throughput optimality for any arrival rate. To state the proof, we first introduce the definition of a *section* and its weight.

***Definition** 1:* We let $q^f_{\max}(t) \triangleq \max_{i\in f} q^f_i(t)$ where $q^f_i(t)$ is the queue length of node i belongs to flow f at time t. A *section* $S^f(t)$ of flow f at time t is a set of consecutive nodes that can be represented as

$$\left(n_1^{(f,s1)}, n_2^{(f,s1)}, ..., n_k^{(f,s1)}, n_1^{(f,s2)}, n_2^{(f,s2)}, ..., n_s^{(f,s2)}\right)$$

where $q^f_{n_j^{(f,s2)}}(t) = q^f_{\max}(t), \forall j \in [1,s]$, $(1-\delta)q^f_{\max}(t) < q^f_{n_j^{(f,s1)}}(t) < q^f_{\max}(t)$, $\forall j \in [1,k]$ given $\delta \in (0,1)$, with s and k representing the number of services of these two types, respectively. The set of *sections* for flow f at period t is denoted as $\mathcal{S}^f(t)$.

***Definition** 2:* The weight of a *section* at time t is defined as the average packages that one node holds in the *section*, i.e.,

$$\omega_{S^f}(q^f(t)) \triangleq \frac{\sum_{n\in S^f(t)} q^f_n(t)}{|S^f(t)|}$$

where $|S^f(t)|$ is the number of nodes in *section* $S^f(t)$.

Specifically, a *section* has the following properties:

(i) the last node of the *section* has the maximum number of packages of the flow, and the number of packages of the next node of the last node, if any, is strictly less than $q^f_{\max}(t)$;

(ii) the number of packages that other nodes possess in the *sections* of the same flow f is strictly larger than $(1-\delta)q^f_{\max}(t)$.

Proof of the Theorem.

For any arrival rate $\lambda_f(t)$, there always exists a real number $\epsilon \in (0,1/2)$ such that $(1+\epsilon)(\sum_{f:f\in\mathcal{F}} \lambda_f(t)) \in \Lambda$. We select a $\delta = \epsilon/2 \in (0,1/4)$. In our proof, we define the following Lyapunov function

$$V(q^f(t)) \triangleq \max_{f\in\mathcal{F}} \frac{1}{\lambda_f(t)} \max_{S^f(t)\in\mathcal{S}^f(t)} \omega_{S^f(t)}(q^f(t)), \quad (1)$$

where $q^f(t)$ is the queue length of f at time t for $f \in \mathcal{F}$. We would like to show $\frac{D^+}{dt^+}V(q^f(t)) \le 0$ whenever $V(q^f(t)) > 0$ according to Lemma 1:

***Lemma** 1:* Let $g : [0,\infty) \to [0,\infty)$ be a locally Lipschitz continuous function.

(i) Assume that $g(0) = 0$ and $\frac{D^+}{dt^+}g(t) \le 0$ whenever $g(t) > 0$. Then, $g(t) = 0$ for all $t \ge 0$.

(ii) Assume that $g(0) > 0$ and $\frac{D^+}{dt^+}g(t) \le -\gamma$ for some $\gamma > 0$ whenever $g(t) > 0$. Then, there exists a $T \ge 0$ such that $g(t) = 0$ for all $t \ge T$.

This lemma implies that the local Lipschitz continuous function will converge to zero under certain specific conditions.

For the derivation of a function $f(x)$, we have the following lemma:

***Lemma** 2:* We define the right derivative of f as

$$\frac{D^+}{dx^+} f(x) \triangleq \limsup_{u\to 0} \frac{f(x+u) - f(x)}{u}$$

and

$$\mathcal{K} \triangleq \{i | f_i(x) = f(x)\}.$$

If $f(x) = \max_{i=1,2,,...,K} f_i(x)$ and $f_i(x)$, $\forall i$, are locally Lipschitz continuous, then we have

$$\frac{D^+}{dx^+} f(x) \le \max_{i\in\mathcal{K}} \left\{ \frac{D^+}{dx^+} f_i(x) \right\}.$$

This lemma extends the assumption of the differentiability of $f_i(x)$ to encompass scenarios where these functions exhibit local Lipschitz continuity.

Utilizing Lemma 2, we can obtain

$$\frac{D^+}{dt^+} V(q^f(t)) \le \max_{\overline{f}\in\overline{\mathcal{K}}} \frac{1}{\lambda_{\overline{f}}(t)} \frac{D^+}{dt^+} \max_{S^{\overline{f}}(t)\in\mathcal{S}^{\overline{f}}(t)} \omega_{S^{\overline{f}}(t)}(q^f(t))$$

where

$$\overline{\mathcal{K}}(t) \triangleq \left\{ \overline{f} \in \mathcal{F} : V(q^f(t)) = \frac{1}{\lambda_{\overline{f}}(t)} \max_{S^{\overline{f}}(t)\in\mathcal{S}^{\overline{f}}(t)} \omega_{S^{\overline{f}}(t)}(q^f(t)) \right\}.$$

Next, we consider the case when $V(q^f(t)) > 0$, i.e.,

$$\max_{S^{\overline{f}}(t)\in\mathcal{S}^{\overline{f}}(t)} \omega_{S^{\overline{f}}(t)}(q^f(t)) > 0.$$

Here we present a lemma used in our demonstration:

***Lemma** 3:* Assume $\theta(\sum_{f:f\in\mathcal{F}} \lambda_f) \in \Lambda$ for some $\theta > 0$. If $q^f_i \ne 0$ and $(i^*, f^*) \in \arg\max_{(i,f)} \frac{q^f_i}{\lambda_f}$, then $\mu^{f^*}_{i^*} \ge \theta\lambda_{f^*}$.

This lemma indicates that the pair (i^*, f^*) maximizes $\frac{q^f_i}{\lambda_f}$ over all pairs (i,f) on the premise of $\theta(\sum_{f:f\in\mathcal{F}} \lambda_f) \in \Lambda$ for some $\theta > 0$. And the departure rate of f^* at node i^* is at least $\theta\lambda_{f^*}$ under QueueFlower.

According to Lemma 3, we have

$$\frac{q^{f^*}_{i^*}(t)}{\lambda_{f^*}(t)} \ge \frac{q^{\overline{f}}_{\max}(t)}{\lambda_{\overline{f}}(t)} \overset{(a)}{\ge} \frac{1}{\lambda_{\overline{f}}(t)} \max_{S^{\overline{f}}(t)\in\mathcal{S}^{\overline{f}}(t)} \omega_{S^{\overline{f}}(t)}(q^f(t)) > 0,$$

where the step (a) follows from the property of a *section* $S^f(t)$ that the weight of any *section* of flow $\underline{f}$ must not exceed $q^f_{\max}(t)$. Hence we have $q^{f^*}_{i^*}(t) > 0$ and $q^{\overline{f}}_{\max}(t) > 0$.

Now we consider $\frac{D^+}{dt^+} \max_{S^{\overline{f}}(t)\in\mathcal{S}^{\overline{f}}(t)} \omega_{S^{\overline{f}}(t)}(q^f(t))$. Since $q^{\overline{f}}_i(t)$ is continuous and $q^{\overline{f}}_{\max}(t) > 0$, $q^{\overline{f}}_i(t)$ is equipped with sign-preserving property for a sufficiently small $u > 0$. So we have two facts:

(i) if a node has the most number of packages of $S^{\overline{f}}$ at time $t+u$, it should also have the most number of packages at time t;

(ii) the node with the number of flow $\overline{f}$ packages $q^{\overline{f}}_i(t) \leqslant (1-\delta)q^{\overline{f}}_{\max}(t)$ should also has the number of flow $\overline{f}$ packages $q^{\overline{f}}_i(t+u) \leqslant (1-\delta)q^{\overline{f}}_{\max}(t+u)$.

We assume a part of *sections* of $\overline{f}$ at time $t+u$ is

$$\left(n_1^{(\overline{f},s0)},...,n_m^{(\overline{f},s0)},n_1^{(\overline{f},s1)},...,n_k^{(\overline{f},s1)},n_1^{(\overline{f},s2)},...,n_s^{(\overline{f},s2)}\right),$$

where s is the number of nodes with the maximum queue length in $S^{\overline{f}}(t+u)$, and $m+k$ is the number of nodes whose queue length is less than $q^{\overline{f}}_{\max}(t+u)$ in $S^{\overline{f}}(t+u)$.

By fact(ii), we have

$$q^{\overline{f}}_{n_I^{(\overline{f},s0)},\overline{f}}(t+u) < min\{q^{\overline{f}}_{n_J^{(\overline{f},s1)},\overline{f}}(t+u), q^{\overline{f}}_{n_K^{(\overline{f},s2)},\overline{f}}(t+u)\}.$$

So we have

$$\begin{aligned}&\max_{S^{\overline{f}}(t+u)\in\mathcal{S}^{\overline{f}}(t+u)}\omega_{S^{\overline{f}}(t+u)}(q^f(t+u))\\ \leq&\frac{1}{k+s}\left(\sum_{j=1}^{k}q^{\overline{f}}_{n_j^{(\overline{f},s1)}}(t+u)+\sum_{j=1}^{s}q^{\overline{f}}_{n_j^{(\overline{f},s2)}}(t+u)\right).\end{aligned} \tag{2}$$

In addition, since $q^{\overline{f}}_{n_i^{(\overline{f},s1)}}(t) < q^{\overline{f}}_{n_j^{(\overline{f},s2)}}(t), \forall i,j$, we have

$$\begin{aligned}&\max_{S^{\overline{f}}(t)\in\mathcal{S}^{\overline{f}}(t)}\omega_{S^{\overline{f}}(t)}(q^f(t))\\ \geq&\frac{1}{k+s}\left(\sum_{j=1}^{k}q^{\overline{f}}_{n_j^{(\overline{f},s1)}}(t)+\sum_{j=1}^{s}q^{\overline{f}}_{n_j^{(\overline{f},s2)}}(t)\right).\end{aligned} \tag{3}$$

By utilizing (2) and (3), we have

$$\begin{aligned}&\frac{D^+}{dt^+}\max_{S^{\overline{f}}(t)\in\mathcal{S}^{\overline{f}}(t)}\omega_{S^{\overline{f}}(t)}(q^f(t))\\ =&\limsup_{u\to0}\frac{1}{u}\big(\max_{S^{\overline{f}}(t+u)\in\mathcal{S}^{\overline{f}}(t+u)}\omega_{S^{\overline{f}}(t+u)}(q^f(t+u))\\ &-\max_{S^{\overline{f}}(t)\in\mathcal{S}^{\overline{f}}(t)}\omega_{S^{\overline{f}}(t)}(q^f(t))\big)\\ \leq&\limsup_{u\to0}\frac{1}{k+s}\big(\sum_{j=1}^{k}\frac{1}{u}(q^{\overline{f}}_{n_j^{(\overline{f},s1)}}(t+u)-q^{\overline{f}}_{n_j^{(\overline{f},s1)}}(t))\\ &+\sum_{j=1}^{s}\frac{1}{u}(q^{\overline{f}}_{n_j^{(\overline{f},s2)}}(t+u)-q^{\overline{f}}_{n_j^{(\overline{f},s2)}}(t))\big)\\ =&\frac{1}{k+s}\big(\sum_{j=1}^{k}\frac{d}{dt}q^{\overline{f}}_{n_j^{(\overline{f},s1)}}(t)+\sum_{j=1}^{s}\frac{d}{dt}q^{\overline{f}}_{n_j^{(\overline{f},s2)}}(t)\big).\end{aligned}$$

Now we discuss two cases regarding whether $n_1^{(\overline{f},s1)}$ is the frontend. We omit the time index t for concision.

(i) $n_1^{(\overline{f},s1)}$ is not the frontend for flow $\overline{f}$.

We let $n_{j\_}$ denote the previous node of j-th node, then $n_{1\_}^{(\overline{f},s1)}$ represents the previous node of the first node of s_1 of flow f, i.e., the previous node of *section* $S^{\overline{f}}$.

We have

$$\begin{aligned}&\frac{1}{\lambda_{\overline{f}}}\frac{D^+}{dt^+}\max_{S^{\overline{f}}\in\mathcal{S}^{\overline{f}}}\omega_{S^{\overline{f}}}(q^f)\\ \leq&\frac{1}{\lambda_{\overline{f}}(k+s)}\left(\sum_{j=1}^{k}(\mu^{\overline{f}}_{n_{j\_}^{(\overline{f},s1)}}-\mu^{\overline{f}}_{n_j^{(\overline{f},s1)}})\right)\\ &+\left(\sum_{j=1}^{s}(\mu^{\overline{f}}_{n_{j\_}^{(\overline{f},s2)}}-\mu^{\overline{f}}_{n_j^{(\overline{f},s2)}})\right)\\ =&\frac{1}{\lambda_{\overline{f}}(k+s)}\left(\mu^{\overline{f}}_{n_{1\_}^{(\overline{f},s1)}}-\mu^{\overline{f}}_{n_s^{(\overline{f},s2)}}\right).\end{aligned} \tag{4}$$

If $q^{\overline{f}}_{n_{1\_}^{(\overline{f},k)}}=0$, then $\mu^{\overline{f}}_{n_{1\_}^{(\overline{f},k)}}=0$ and we have

$$\frac{1}{\lambda_{\overline{f}}}\frac{D^+}{dt^+}\max_{S^{\overline{f}}\in\mathcal{S}^{\overline{f}}}\omega_{S^{\overline{f}}}(q^f)\leq-\frac{\mu^{\overline{f}}_{n_s^{(\overline{f},s2)}}}{\lambda_{\overline{f}}(k+s)}. \tag{5}$$

According to Lemma 3, we can infer that

$$\begin{aligned}\frac{q^{\overline{f}}_{\max}}{\lambda_{\overline{f}}}&\overset{(a)}{\geq}\frac{1}{\lambda_{\overline{f}}}\max_{S^{\overline{f}}\in\mathcal{S}^{\overline{f}}}\omega_{S^{\overline{f}}}(q^f)\\ &\overset{(b)}{\geq}\frac{1}{\lambda_{f^*}}\max_{S^{f^*}\in\mathcal{S}^{f^*}}\omega_{S^{f^*}}(q^f)\\ &\overset{(c)}{\geq}\frac{(1-\delta)q^{f^*}_{i^*}}{\lambda_{f^*}}.\end{aligned} \tag{6}$$

where the steps (a) and (c) follows from the properties of *section*, step (b) follows from the definition of $\overline{f}$.

Recall that $(i^*,f^*)\in\arg\max_{(i,f)}\frac{q_i^f}{\lambda_f}$ and $\delta=\epsilon/2$, we can obtain

$$\mu^{\overline{f}}_{n_s^{(\overline{f},s2)}}=\frac{q^{\overline{f}}_{n_s^{(\overline{f},s2)}}}{q^{f^*}_{i^*}}\mu^{f^*}_{i^*}=\frac{q^{\overline{f}}_{\max}/\lambda_{\overline{f}}}{q^{f^*}_{i^*}/\lambda_{f^*}}\frac{\lambda_{\overline{f}}}{\lambda_{f^*}}\mu^{f^*}_{i^*}\geq(1+\frac{\epsilon}{4})\lambda_{\overline{f}}. \tag{7}$$

By substituting inequality (7) into (5), we have

$$\frac{1}{\lambda_{\overline{f}}}\frac{D^+}{dt^+}\max_{S^{\overline{f}}\in\mathcal{S}^{\overline{f}}}\omega_{S^{\overline{f}}}(q^f)\leq-\frac{1+\epsilon/4}{k+s}\leq-\frac{1+\epsilon/4}{N}$$

due to the fact that $k+s\leq N$.

The resource allocation mechanism for flow is the same as that for nodes under QueueFlower. If $q^{\overline{f}}_{n_{1\_}^{(\overline{f},s1)}}>0$, inequality (4) can be written as

$$\begin{aligned}&\frac{1}{\lambda_{\overline{f}}}\frac{D^+}{dt^+}\max_{S^{\overline{f}}\in\mathcal{S}^{\overline{f}}}\omega_{S^{\overline{f}}}(q^f)\\ \leq&\frac{1}{\lambda_{\overline{f}}(k+s)}\left(\frac{q^{\overline{f}}_{n_{1\_}^{(\overline{f},s1)}}}{q_{n_{1\_}}}\mu_{n_{1\_}}-\frac{q^{\overline{f}}_{n_s^{(\overline{f},s2)}}}{q_{n_s}}\mu_{n_s}\right)\\ \overset{a}{=}&\frac{1}{\lambda_{\overline{f}}(k+s)}(q^{\overline{f}}_{n_{1\_}^{(\overline{f},s1)}}-q^{\overline{f}}_{n_s^{(\overline{f},s2)}})\frac{\mu_{n_s}}{q_{n_s}}\\ \overset{b}{\leq}&\frac{-\delta q^{\overline{f}}_{\max}}{\lambda_{\overline{f}}(k+s)}\frac{\mu}{Q},\end{aligned} \tag{8}$$

where $\mu_{n_s} = \sum_{f:n_s\in f} \mu_{n_s}^f$, $q_{n_s} = \sum_{f:n_s\in f} q_{n_s}^f$, $\mu = \sum_{i\in\mathcal{N}} \mu_i$, $Q = \sum_{i\in\mathcal{N}} q_i$. The step (a) follows that $\frac{\mu_{n_{1\_}}}{q_{n_{1\_}}} = \frac{\mu_{n_s}}{q_{n_s}}$, step (b) follows that $\frac{x_i}{y_i} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i}$ if $\frac{x_1}{y_1} = \frac{x_2}{y_2} = \cdots = \frac{x_n}{y_n}$.

Since (i^*, f^*) maximizes $\frac{q_i^f}{\lambda_f}$, we have $\frac{q_{i^*}^{f^*}}{\lambda_{f^*}} \geq \frac{q_i^f}{\lambda_f}$. Considering the capacity region, we have

$$q_i = \sum_{f:i\in f} q_i^f \leq \frac{q_{i^*}^{f^*}}{\lambda_{f^*}} \sum_{f:i\in f} \lambda_f \leq \mu_i \frac{q_{i^*}^{f^*}}{\lambda_{f^*}}, \tag{9}$$

where μ_i is the capacity of node i.

So we have

$$Q = \sum_{i\in\mathcal{N}} q_i \leq N \frac{q_{i^*}^{f^*}}{\lambda_{f^*}}. \tag{10}$$

Combining (6) and (10), we have

$$\frac{q_{\max}^{\overline{f}}}{\lambda_{\overline{f}}(k+s)} \frac{\mu}{Q} \geq \frac{(1-\delta)\mu}{N(k+s)} \geq \frac{(1-\delta)\mu}{N^2}. \tag{11}$$

Substitute (11) into (8) and recall the fact that $\delta \in (0, 1/4)$, we have

$$\frac{1}{\lambda_{\overline{f}}} \frac{D^+}{dt^+} \max_{S^{\overline{f}}\in\mathcal{S}^{\overline{f}}} \omega_{S^{\overline{f}}}(q^f) \leq -\frac{\delta(1-\delta)\mu}{N^2},$$

which is strictly negative.

(ii) $n_1^{(\overline{f},s1)}$ is the frontend for flow $\overline{f}$ at time t.

Utilizing (4), we can derive

$$\begin{aligned} \frac{1}{\lambda_{\overline{f}}} \frac{D^+}{dt^+} \max_{S^{\overline{f}}\in\mathcal{S}^{\overline{f}}} \omega_{S^{\overline{f}}}(q^f) &\leq \frac{1}{\lambda_{\overline{f}}(k+s)} (\lambda_{\overline{f}} - \mu_{n_s^{(\overline{f},s2)}}^{\overline{f}}) \\ &\overset{(a)}{\leq} -\frac{\epsilon}{4(k+s)} \\ &\overset{(b)}{\leq} -\frac{\epsilon}{4N} \end{aligned} \tag{12}$$

from (4) and the step (a) follows (7), (b) is true because the *section* is part of the nodes of the system.

Therefore, $\frac{D^+}{dt^+} V(q^f(t)) \leq 0$ always holds under a fine-tuned weight w, which implies that QueueFlower achieves throughput optimality for any arrival rate vector within the maximum throughput region.

Proof of Lemma 1.

(i) We assume that there exist a $t_1 > 0$ and a $\zeta > 0$ such that $g(t_1) = \zeta$. Since $g(0) = 0$, there exists a $t_2 \in (0, t_1)$ such that $g(t_2) = \zeta/2$ and $g(t) \geq \zeta/2 > 0$ for any $t \in (t_2, t_1]$. Since we have $g(t) > 0$ for any $t \in [t_2, t_1]$ and $\frac{D^+}{dt^+} g(t) \leq 0$ whenever $g(t) > 0$, we have $g(t_1) \leq g(t_2)$, which contradicts that $g(t_1) = \zeta > g(t_2) = \zeta/2$. Therefore, $g(t) = 0$ for all $t \geq 0$.

(ii) Since $\frac{D^+}{dt^+} g(t) \leq -\gamma$ whenever $g(t) > 0$, the function $g(\cdot)$ will first hit zero at time $T = g(0)/\gamma$. Then, by using the technique in (i), we can show that $g(t) = 0$ for all $t \geq T$.

Proof of Lemma 2.

We prove Lemma 1 by contradiction. We assume

$$\frac{D^+}{dt^+} f(x) > \max_{i\in\mathcal{K}} \left\{ \frac{D^+}{dt^+} f_i(x) \right\}. \tag{13}$$

For a sufficient small ρ, there exists a decreasing sequence $\{u_k, k = 1, 2, ...\}$ with $\lim_{k\to\infty} u_k = 0$ such that

$$\frac{f(x+u_k) - f(x)}{u_k} \geq \max_{i\in\mathcal{K}} \left\{ \frac{D^+}{dt^+} f_i(x) \right\} + \rho, \forall k = 1, 2, \cdots.$$

Note that $f(x) = f_i(x), \forall i \in \mathcal{K}$. Since there are a finite number of local Lipschitz continuous functions $f_i(x), i = 1, 2, \cdots, K$, there must exist a $j \in \mathcal{K}$ and a decreasing subsequence $\{u_{t_k}, k = 1, 2, \cdots\}$ of $\{u_k, k = 1, 2, \cdots\}$ such that $f_j(x + u_{t_k}) = f(x + u_{t_k}) = \max_{i=1,2,\cdots,K} f_i(x + u_{t_k}), \forall k = 1, 2, \cdots$, which implies that

$$\frac{f_j(x+u_{t_k}) - f_j(x)}{u_{t_k}} \geq \max_{i\in\mathcal{K}} \left\{ \frac{D^+}{dt^+} f_i(x) \right\} + \rho, \forall k = 1, 2, \cdots.$$

Therefore, we have the contradiction

$$\frac{D^+}{dt^+} f_j(x) \geq \max_{i\in\mathcal{K}} \left\{ \frac{D^+}{dt^+} f_i(x) \right\} + \rho. \tag{14}$$

So we have the desired result.

Proof of Lemma 3.

We have $q_{i^*}^{f^*} > 0$. Assume $\mu_{i^*}^{f^*} < \theta\lambda_{f^*}$. Then for any flow within the node $i (f \neq f^*)$, there are two cases:

(i) If $q_i^f = 0$, then $\mu_i = 0$.

(ii) If $q_i^f > 0$, then we have

$$\mu_i^f \overset{(a)}{=} \frac{q_i^f}{q_{i^*}^{f^*}} \mu_{i^*}^{f^*} = \frac{q_i^f/\lambda_f}{q_{i^*}^{f^*}/\lambda_{f^*}} \frac{\lambda_f}{\lambda_{f^*}} \mu_{i^*}^{f^*} \overset{(b)}{\leq} \lambda_f \frac{q_{i^*}^{f^*}}{\lambda_{f^*}} \overset{(c)}{<} \theta\lambda_f$$

where the step (a) follows the resource allocation mechanism of QueueFlower, (b) is true since $(i^*, f^*) \in \arg\max_{(i,f)} \frac{q_i^f}{\lambda_f}$, (c) follows from our assumption.

Combining (i) and (ii), we have $\mu_i^f < \theta\lambda_f$ in any case. Hence we have $\mu < \theta \sum_{\lambda_f\in\Lambda} \lambda_f$, which contradicts the fact that $\sum_{f:n\in f} \kappa(f,n) \cdot \lambda_f)_{n\in\mathcal{N}} \in \Lambda$, i.e., the service rate vector strictly lies on the boundary of the capacity region Λ. Therefore, we have the desired result.