Proofs of the Propositions, Lemmas and Theorems

Proof of Proposition 1. Since $\mathcal{C}$ is a general server, it satisfies condition (a) of Definition 3, i.e., $\lim_{T\to\infty} m(\{t|\prod_{k\in\mathcal{C}} \dot{R}_k(t) \neq 0, t \leq T\})/m(\{t|\sum_{k\in\mathcal{C}} \dot{R}_k(t) \neq 0, t \leq T\}) = 0$. Following from $\mathcal{C}' \supseteq \mathcal{C}$, we also have $\{t|\prod_{k\in\mathcal{C}} \dot{R}_k(t) \neq 0, t \leq T\} \supseteq \{t|\prod_{k\in\mathcal{C}'} \dot{R}_k(t) \neq 0, t \leq T\}$ and $\{t|\sum_{k\in\mathcal{C}} \dot{R}_k(t) \neq 0, t \leq T\} \subseteq \{t|\sum_{k\in\mathcal{C}'} \dot{R}_k(t) \neq 0, t \leq T\}$. Therefore, there is $\lim_{T\to\infty} m(\{t|\prod_{k\in\mathcal{C}} \dot{R}_k(t) \neq 0, t \leq T\})/ m(\{t|\sum_{k\in\mathcal{C}} \dot{R}_k(t) \neq 0, t \leq T\}) = 0$. Together with condition (b), it follows that $\mathcal{C}'$ is also a general server.

Proof of Lemma 1. Consider any $\alpha \in [0.8, 0.9)$ and assume the 1st job arrives at class 1 at $t = 0$. Since the interarrival times are constant, the state $X(t)$ of the system at time t reduces to a two-tuple $X(t) = (Q(t), V(t))$. At $t = 0$, $X(0) = ((0, 0.2),\ (0, 0), (0,0), (0,0))$. At $t = \alpha$, the 1st job is still at class 3, and therefore $X(\alpha) = ((0, 0.2),\ (0, 0), (0, 0.9 - \alpha), (0, 0))$. At $t = 0.9$, the 1st job arrives at class 4 but needs to wait since the 2nd job is still receiving service at class 1. Hence, $X(0.9) = ((0, \alpha - 0.7), (0, 0), (0, 0), (1, 0))$. At $t = 2\alpha$, the 3rd job arrives, the 1st job has leaved the network, and the 2nd job is at class 3. Thus, $X(2\alpha) = ((0, 0.2),$ $(0, 0), (0, 0.9 - \alpha), (0, 0))$. Note that the state of the network at time 2α is identical to the state at time α. Since the interarrival times and services times are both constant, $X(n\alpha) = ((0, 0.2),\ (0, 0), (0, 0.9 - \alpha), (0, 0))$ for any integer $n \geq 1$ and the evolutions in the interval $((n - 1)\alpha, n\alpha)$ are the same. In addition, the time epoch that the nth job begins to receive service at class 2 is $n\alpha + 0.2$, and so is the time epoch that the $(n - 1)$th job begins to receive service at class 4. Therefore, except for the first job to initiate the system, classes 2 and 4 always process jobs simultaneously. Q.E.D.

Proof of Lemma 2. Let the arrival and departure times of the nth job at class i be A_n^i and D_n^i. Assume there are 3 jobs between classes 2 and 4, i.e., the nth, $(n + 1)$th and $(n + 2)$th jobs. There can be the following situations:

(a) The $(n + 1)$th job at class 2 and the nth job at class 4 are processed simultaneously and the $(n + 2)$th job is waiting at class 2. Since the $(n + 2)$th job will block the $(n + 1)$th job, there is no job receiving service at class 4 while the $(n + 2)$th job is receiving service at class 2, i.e., class 4 is vacant during the interval (A_{n+2}^2, D_{n+2}^2). Therefore, the synchronization breaks.

(b) The $(n + 2)$th job at class 2 and the nth job at class 4 are processed simultaneously and the $(n + 1)$th job is waiting at either class 3. After D_{n+2}^2, the $(n + 1)$th job will arrive at class 4 at time $D_{n+2}^2 + 0.1$ and A_{n+2}^4 will be $D_{n+2}^2 + 0.2$. Since at most one job (i.e., the $(n + 3)$th job) can be at class 2 at time $D_{n+2}^2 + 0.2$, classes 2 and 4 cannot process jobs simultaneously at time $D_{n+2}^2 + 0.8$ and the synchronization breaks.

(c) The $(n + 2)$th job at class 2 and the nth job at class 4 are processed simultaneously and the $(n + 1)$th job waits at class 4. the $(n + 1)$th job will block class 1 after the nth job completes its service, i.e., D_n^4 . Hence, classes 2 and 4 cannot process jobs simultaneously and the synchronization breaks.

Similar arguments can be applied when there are more than three jobs between classes 2 and 4. Q.E.D.

Proof of Lemma 3. Since the network is synchronized when $0.8 \leq \alpha < 0.9$ and $1.1 \leq \alpha < 1.2$, we will evaluate if the network is synchronized when $0.9 \leq \alpha < 1.1$.

By Lemma 2, there are at most two jobs between class 2 and class 4 in the Lu-Kumar network when $\{2, 4\}$ is synchronized. Classes 2 and 4 are synchronized if and only if both always are busy at the same time (except for the first job to initialize the system). Hence, if the network is synchronized, the relations of the two jobs can be one of the following two:

(a) One job may wait at class 4: $A_n^4 \le D_{n+1}^1$ (two consecutive jobs can be served at the same time), $A_{n+1}^1 < A_n^4$, $A_{n+1}^2 \ge D_n^3$ and $A_n^3 < A_{n+1}^2$ (no consecutive job can be blocked at classes 1, 2 and 3), or

(b) One job may wait at class 2: $A_{n+1}^2 \le D_n^3$ (two consecutive jobs can be served at the same time), and $A_{n+1}^1 < A_n^4$, $A_n^3 < A_{n+1}^2$ and $A_{n+1}^4 \ge D_n^1$ (no consecutive job can be blocked at classes 1, 3 and 4).

For (a), since $A_{n+1}^1 < A_n^4 \le D_{n+1}^1$, $A_{n+1}^1 < A_n^4 \le A_{n+1}^1 + 0.2$, we have $0 < A_n^4 - A_{n+1}^1 \le 0.2$. Because no consecutive job can be blocked at classes 1, 2 and 3, $A_n^4 - A_n^1 = 0.9$. Therefore, $0.7 \le A_{n+1}^1 - A_n^1 < 0.9$, which is a contradiction to $0.9 \le \alpha < 1.1$.

For (b), since $A_n^3 < A_{n+1}^2 \le D_n^3$, $A_n^3 < A_{n+1}^2 \le A_n^3 + 0.1$, we have $0 < A_{n+1}^2 - A_n^3 \le 0.1$. Since $m_2 = 0.6$, $m_3 = 0.1$ and the nth and $(n-1)$th jobs are served at classes 2 and 4 simultaneously, we have $A_n^2 + 0.6 \le A_n^3 < A_n^2 + 0.7$ (if nth job is blocked by the $(n-1)$th job at class 2). Hence, $0.6 < A_{n+1}^2 - A_n^2 < 0.8$. Because no consecutive job can be blocked at class 1, $0.6 < A_{n+1}^1 - A_n^1 < 0.8$, which is a contradiction to $0.9 \le \alpha < 1.1$.

Together with Lemma 1, the set $\{2, 4\}$ (i.e., classes 2 and 4) can be synchronized if and only if $0.8 \le \alpha < 0.9$ or $1.1 \le \alpha < 1.2$. Q.E.D.

Proof of Theorem 1. (i) If a queueing network is stable, then effective traffic intensity of every general server does not exceed one. Proceeding by contradiction, we assume there exists some general server S with M effective classes such that $P_S > 1$. Since $\lim_{t\to\infty} D_k(t)/t = \lambda_{\tau(k)}$ for any $k \in S$ (due to the pathwise stability), we have $\lim_{t\to\infty} T_k(t)/t = \lambda_{\tau(k)} m_k$ ($k \in S$), where $T_k(t)$ is the cumulative service time received by class k. Since $P_S > 1$, we have $\sum_{k\in S} \lim_{t\to\infty} T_k(t)/t = \sum_{k\in S} \lambda_{\tau(k)} m_k > M$. On the other hand, the definition of M implies that $\sum_{k\in S} \lim_{t\to\infty} T_k(t)/t \le M$, which contradicts to our earlier conclusion. Hence, the effective traffic intensity of every general server does not exceed one if the network is stable. (ii) If the effective traffic intensity of every general server does not exceed one, then the queueing network is stable. It suffices to show that the corresponding fluid model is weakly stable if the effective traffic intensity of each general server does not exceed one. For any $1 \le j \le J$ and $1 \le k \le K$, the basic fluid equations are given as follows:

$$
\begin{aligned}
&Q_k(t) = Q_k(0) + A_k(t) - \mu_k T_k(t),\\
&I_j(t) = t - \textstyle\sum_{k:\,\sigma(k)=j} T_k(t),\\
&Q_k(t) \ge 0,\\
&T_k(0) = 0,\\
&I_j(0) = 0,\ I_j(\cdot) \text{ is non-decreasing, and}\\
&\dot{I}_j(t) = 0, \text{ when } \textstyle\sum_{k:\,\sigma(k)=j} Q_k(t) > 0 \text{ and } I_j(t) \text{ is differentiable at } t,
\end{aligned}
$$

where $A_k(t) = \lambda_{\tau(k)} t$ if k is the first class of type $\tau(k)$, otherwise $A_k(t) = \mu_{k-1} T_{k-1}(t)$. Similar to physical stations, for any general server S with M effective servers, we have extra fluid equations: $I_s(t) = Mt - \sum_{k\in S} T_k(t)$ is non-decreasing and $P_S \le 1$. Proceeding by contradiction, we assume there is a fluid allocation which is not weakly stable, i.e., given $Q(0) = 0$, there exists a solution such that $Q(t_0) \ne 0$ for some t_0. Then there exists a smallest k^* and a pair of t_1 and t_2 less than t_0 such that $Q(t) = 0$ for $t \in (0, t_1)$, $Q_{k^*}(t) > 0$ for $t \in (t_1, t_2]$. Since the service policy is work-conserving, for at least one general server S^* which contains k^*, we have $\sum_{k\in S^*}(T_k(t_2) - T_k(t_1)) = EF(S^*)(t_2 - t_1)$. Furthermore, it follows that $\sum_{k\in S^*} \lambda_{\tau(k)}(t_2 - t_1)m_k - EF(S^*)(t_2 - t_1) = \sum_{k\in S^*} \lambda_{\tau(k)}(t_2 - t_1)m_k - \sum_{k\in S^*}(T_k(t_2) - T_k(t_1)) \ge \sum_{k\in S^*} Q_k(t_2) m_k > 0$, which implies that $P_{S^*} > 1$.

This is a contradiction to the given condition. Hence, the fluid model is weakly stable and the network is pathwise stable. Q.E.D.

REMARK: Since general servers are defined for a queueing network under a given service discipline and the structure of general servers also depends on the service discipline, general servers will include the information of the service discipline in some sense. Hence, in this proof, we do not need the extra fluid equation corresponding to the service discipline (e.g. FIFO policy). Considering the equations corresponding to the general servers under the given service discipline is sufficient.

Proof of Theorem 2. Based on Theorem 1, it suffices to show that if the effective traffic intensity of every compact server does not exceed one, it is also the case for general servers. Considering any general server S with M effective classes, we have the following four situations:

(i) S itself is a compact server, then $P_S \le 1$ trivially.

(ii) S is a subset of a compact server, i.e., $S \subseteq S'$, where S' is a compact server with the same effective number of servers M as S and we have $P_{S'} \le 1$. For general server S, we have $P_S = L_S/EF(S) \le L_{S'}/EF(S) = L_{S'}/EF(S') = P_{S'} \le 1$.

(iii) S is the union of compact servers, i.e., $S = \bigcup_{i\in F} S_i$, where each S_i is a compact server, F is an index set and $S_i \cap S_{i'} = 0$ for any $i \ne i'$. We have $EF(S) \ge \sum_{i\in F} EF(S_i)$ because of the compactness of S_i. Then it follows that $P_S = L_S/EF(S) = \sum_{i\in F} L_{S_i}/EF(S) \le \sum_{i\in F} L_{S_i}/\sum_{i\in F} EF(S_i) \le 1$.

(iv) $S = (\bigcup_{i\in F} S_i) \cup S_j$, where each S_i is a compact server, F is an index set, $S_i \cap S_{i'} = 0$ for any $i \ne i'$ and S_j is a subset of a compact server. We first consider $(\bigcup_{i\in F} S_i) \cup S_{j'}$, where $S_{j'}$ is a compact server containing S_j and $EF(S_{j'}) = EF(S_j)$. It follows from (iii) that $P_{(\bigcup_{i\in F} S_i)\cup S_{j'}} \le 1$. Then we have $P_S = L_S/EF(S) \le L_{(\bigcup_{i\in F} S_i)\cup S_{j'}}/EF(S) = L_{(\bigcup_{i\in F} S_i)\cup S_{j'}}/EF((\bigcup_{i\in F} S_i) \cup S_{j'}) \le 1$.

Therefore, a queueing network is stable if and only if the effective traffic intensity of every compact server does not exceed one. Q.E.D.

Proof of Proposition 2. When there is only one station, i.e., $J = 1$, it is trivial that the physical station is the only compact server. Based on Theorem 2, the network is stable since the traffic intensity of the physical station does not exceed one. For a two-station feedforward network, it suffices to show that the effective traffic intensity of any class set $\mathcal{C} = \{c_1, c_2\}$, where $\sigma(c_j) = j$ for $j = 1, 2$, is no greater than one. It is trivial that $P_{\mathcal{C}} \le 1$ if $m_{c_1} + m_{c_2} \le 1/\lambda$. Next we will consider $m_{c_1} + m_{c_2} > 1/\lambda$. Since the first station is stable, we have $\lim_{t\to\infty} T_{c_1}(t)/t = \lambda m_{c_1}$, where for any class k, $T_k(t)$ is the cumulative service time received by class k. Since traffic intensity of the second station is no larger than one, we also have $\lim_{t\to\infty} m\{t | R_{c_2}(t) > 0, \sum_{k\in station\ 2, k\ne c_2} R_k(t) = 0\}/t \ge \lambda m_{c_2}$. Since the service policy is work-conserving, we further have $\lim_{t\to\infty} m\{t | \dot{R}_{c_2}(t) > 0\}/t \ge \lambda m_{c_2}$. Since $m_{c_1} + m_{c_2} > 1/\lambda$ and $\lim_{t\to\infty} m\{t | \dot{R}_{c_1}(t) > 0\}/t = \lambda m_{c_1}$, $\lim_{t\to\infty} m\{t | \dot{R}_{c_1}(t)\dot{R}_{c_2}(t) \ne 0\}/ m\{t | \dot{R}_{c_1}(t) + \dot{R}_{c_2}(t) \ne 0\} > 0$. Hence, $\mathcal{C} = \{c_1, c_2\}$ is not a general server and $P_{\mathcal{C}} = (\lambda m_{c_1} + \lambda m_{c_2})/2 \le 1$. By induction, any feedforward network is stable if the traffic intensity at every physical station does not exceed one. Q.E.D.

Proof of Proposition 3. For a two-station reentrant line, it is sufficient to show that the effective traffic intensity of any class set $\mathcal{C} = \{c_1, c_2\}$, where $\sigma(c_j) = j$ for $j = 1, 2$, is no greater than one. It is trivial that $P_{\mathcal{C}} \le 1$ if $m_{c_1} +$

$m_{c_2} \leq 1/\lambda$. Next we will consider $m_{c_1} + m_{c_2} > 1/\lambda$. Without loss of generality, we assume $c_1 < c_2$. First, we consider $c_1 = 1$. Since class 1 has the highest priority at station 1, we have $\dot{R}_1(t) < 0$ if $R_1(t) > 0$. If $c_2 = 2$, then class 2 has the highest priority at station 2 and $\dot{R}_2(t) < 0$ when $R_2(t) > 0$. Since $m_{c_1} + m_{c_2} > 1/\lambda$, $\lim_{t\to\infty} m\{t|\dot{R}_{c_1}(t)\dot{R}_{c_2}(t) \neq 0\}/ m\{t|\dot{R}_{c_1}(t) + \dot{R}_{c_2}(t) \neq 0\} > 0$. Hence, the set $\mathcal{C}$ is not a general server and $P_{\mathcal{C}} \leq 1$. If $c_2 = 3$ and class 2 belongs to the first station, we still have $\lim_{t\to\infty} m\{t|\dot{R}_{c_1}(t)\dot{R}_{c_2}(t) \neq 0\}/ m\{t|\dot{R}_{c_1}(t) + \dot{R}_{c_2}(t) \neq 0\} > 0$. If $c_2 = 3$ and class 2 belongs to the second station, we have $\lim_{t\to\infty} m\{t|\dot{R}_2(t) < 0\}/t = \lambda m_2$ since class 2 has the highest priority. Hence, $\lim_{t\to\infty} m\{t|R_{c_2}(t) > 0,\ R_2(t) = 0\}/t = \min\{\lambda m_{c_2},\ 1 - \lambda m_2\} = \lambda m_{c_2}$, which immediately implies $\lim_{t\to\infty} m\{t|\dot{R}_{c_1}(t)\dot{R}_{c_2}(t) \neq 0\}/ m\{t|\dot{R}_{c_1}(t) + \dot{R}_{c_2}(t) \neq 0\} > 0$. Therefore, $P_{\mathcal{C}}$ is also no greater than one when $c_1 = 1$ and $c_2 = 3$. When $c_2 > 3$, the proof is similar. Hence, the set $\{c_1, c_2\}$ is not a general server and $P_{\mathcal{C}} \leq 1$ when $c_1 = 1$.

When $c_1 = 2$, the arguments are similar because class 1 has the highest priority and we can assume the external arrivals are at class 2. By induction, a two-station reentrant line is stable if the traffic intensity at every physical station does not exceed one. For an J-station reentrant line, the similar analysis applies. Therefore, any reentrant line operating under FBFS discipline is stable if the traffic intensity at every physical station does not exceed one. Q.E.D.

Proof of Theorem 3. It suffices to show that the mutual blocking (of classes belonging to different physical stations) will break if the dispatching policy is WIP-dependent, i.e., class k will be assigned the highest priority if $Q_k > \varphi_k$, where Q_k is the number of jobs in buffer k and φ_k is a preset threshold. Assume there exists a set S (not a physical station) such that $\lim_{t\to\infty} m(\{t|\prod_{k\in S}\dot{R}_k(t) \neq 0\})/ m(\{t|\sum_{k\in S}\dot{R}_k(t) \neq 0\}) = 0$ (S suffers mutual blocking), $P_S > 1$ and $\sum_{k\in S} Q_k(t)/t \nrightarrow 0$. If S only consists of two classes k and l, then $(Q_k(t) + Q_l(t))/t \nrightarrow 0$. Without loss of generality, we assume k is a class before l. For any large enough M, there exists a t_0 s.t. $Q_k(t_0) > M$. Therefore, we have $Q_k(t) \geq \varphi_k$ and $\dot{R}_k(t) < 0$ for any $t \in (t_0, t_0 + T)$, where T is a positive number and $T/(M - \varphi_k) \xrightarrow{a.s.} m_k$ as $M \to \infty$. If l is the next class of k (i.e., $l = k + 1$), Q_l will increase by $(M - \varphi_k)$ during the interval $(t_0, t_0 + T)$ since class l does not receive service. Hence, Q_l will exceed φ_l during the interval $(t_0, t_0 + T)$ and should have been assigned the highest priority. This contradicts to the previous assumption.

If $l = k + 2$, $(Q_{k+1} + Q_l)$ will increase by $(M - \varphi_k)$ during the interval $(t_0, t_0 + T)$. As long as Q_{k+1} exceeds φ_{k+1}, class $k + 1$ will receive service and Q_l increases. Since M is large enough, Q_l will exceed its threshold φ_l during the interval $(t_0, t_0 + T)$, which is contradicted to the previous assumption.

By induction, Q_l will always exceed φ_l in the interval $(t_0, t_0 + T)$ since M is large enough and l is finite. Hence, classes k and l will both be assigned the highest priority during a period, which is a contradiction.

Similar arguments can be applied to the case that S consists of multiple classes. Therefore, a queueing network which satisfies the usual traffic condition will always be stable under the WIP-dependent policy. Q.E.D.