LOWER BOUND AND OPTIMALITY IN SWITCHED NETWORKS

BY DEVAVRAT SHAH* AND DAMON WISCHIK†

MIT and UCL

We consider a queueing network in which there are constraints on which queues may be served simultaneously. Such networks, called “switched networks” [3] can be used to model input-queued switches, wireless networks, or bandwidth sharing in the Internet. The scheduling algorithm for such a network specifies which queues to serve at any point in time. The performance of scheduling algorithm is determined by the induced net queue-size. The question of designing optimal scheduling algorithm with this performance metric has been of great recent interest (e.g. [1, 3, 4]).

An important step in this quest is that of finding fundamental limitations of scheduling algorithms in terms of the induced queue-size. In this paper, we present a novel technique to characterize lower bound on average queue-size induced by any algorithm. Through an example, we establish the tightness of this technique for a class of problems.

1. Setup.

1.1. An abstract model. Our network is a collection of \( n \) queues. Each queue has a dedicated exogeneous arrival process through which new work arrives in the form of unit sized packets. Each queue can be potentially serviced at unit rate resulting in departures of packets from it upon completion of their unit service requirement. The network will be assumed to be single-hop, i.e. once work leaves a queue, it leaves the network. At the first glance, this appears to be a strong limitation. However, the results of this paper in terms of lower bounds can be extended for multi-hop networks and this will be discussed in the longer version of this work.

Let \( t \in \mathbb{R}_+ \) denote the continous time and \( \tau = [t] \in \mathbb{N} \) denote the corresponding discrete time slot. Let \( Q_i(t) \in \mathbb{N} \) be number of packets in the \( i \)th queue at time \( t \). And, define \( Q_i(\tau) = Q_i(\tau^+) \), i.e. the queue-size measured in the very beginning of the time slot \( \tau \). Let \( Q(t), Q(\tau) \) denote the vector of queue sizes \( [Q_i(t)]_{1 \leq i \leq n}, [Q_i(\tau)]_{1 \leq i \leq n} \) respectively. Initially, time \( t = \tau = 0 \) and the system starts empty. That is, \( Q(0) = 0 = [0] \).

*Supported by NSF CNS-0546590 and NSF TF-0728554.
†Supported by a Royal Society university research fellowship. Collaboration supported by the British Council Researcher Exchange program.
Arrival process is assumed to be continuous time. Specifically, arrivals happen in terms of packets, each requiring unit amount of service. Let $A_i(t)$ denote the cumulative arrival process for queue $i$ in time interval $[0, t]$. For concreteness of results, we will assume that the arrival processes are independent across queues and $A_i(\cdot)$ is a Poisson process with rate $\lambda_i$ for each $i$. That is, $\mathbb{E}[A_i(t) - A_i(t - 1)] = \lambda_i$ for all $i$ and $t$. Denote the arrival rate vector as $\lambda = [\lambda_i]_{1 \leq i \leq n}$.

The queues are offered service as per scheduling constraint. The scheduling decisions are made as per a discrete time process. Specifically, let $\sigma(\tau) = [\sigma_i(\tau)]$ denote the scheduling decision in the beginning of time-slot $\tau \in \mathbb{N}$. We assume that the service decision is binary, i.e. $\sigma_i(\tau) \in \{0, 1\}$. If $\sigma_i(\tau) = 1$ and $Q_i(\tau) > 0$ (which is integral) then a packet from the $i$th queue departs instantly. Thus, schedule at any time corresponds to a selection of subset of $n$ queues from which one packet can depart. The scheduling constraints require that only certain subsets of queues can be chosen to be served at unit rate at each time. Let $\mathcal{S} \subset \{0, 1\}^n$ denote the set of all feasible schedules or the set of all simultaneously scheduable queues. Under thus described setup, the schedule $\sigma(\tau)$ at time $\tau$ is such that $\sigma(\tau) \in \mathcal{S} \subset \{0, 1\}^n$.

The queuing dynamics induced under the above described model can be summarized by the following equation: for any $\tau \in \mathbb{N}$ and $1 \leq i \leq n$,

$$Q_i(\tau + 1) = Q_i(\tau) - \sigma_i(\tau)1_{\{Q_i(\tau) > 0\}} + A_i(\tau, \tau + 1),$$

where $A_i(\tau, \tau + 1)$ denotes the total number of arrivals to queue $i$ in time interval $(\tau, \tau + 1]$ and $1_{\{\cdot\}}$ denotes the indicator function. Finally, define the cumulative departure process $D(\tau) = [D_i(\tau)]$, where

$$D_i(\tau) = \sum_{s=0}^{\tau} \sigma_i(s)1_{\{Q_i(s) > 0\}}.$$ 

1.2. An example: Wireless network. We will consider a special case of “switched network” model to quantify the strength of our lower bounding technique. This example models multiple-access wireless single hop network. To this end, consider a network of wireless transmission capcable $n$ devices with the queue $Q_i(\cdot)$ hosted at the device or node $i$. Under any reasonable model of communication deployed in practice (e.g. 802.11 standards), in essence if two devices are close to each other and share a common frequency to transmit at the same time, there will be interference and data is likely to be lost. If the devices are far away, they may be able to simultaneously transmit with no interference. Thus the scheduling constraint here is that no two devices that might interfere with each other can transmit at the same time. This can be naturally modeled as an independent set constraint on a graph (called the interference graph), whose vertices correspond to the devices, and two vertices share an edge if and only if the corresponding devices would interfere. Specifically,
let $G = (V, E)$ be this network interference graph with $V = \{1, \ldots, n\}$ representing $n$ nodes and
$$E = \{(i, j) : i \text{ and } j \text{ interfere with each other}\}.$$ 
Let $\mathcal{N}(i) = \{j \in V : (i, j) \in E\}$ denote the neighbors of node $i$. Let $\mathcal{I}(G)$ denote the set of all independent sets of $G$, i.e. subsets of $V$ so that no two neighbors are adjacent to each other. Formally,
$$\mathcal{I}(G) = \{\sigma = [\sigma_i] \in \{0, 1\}^n : \sigma_i + \sigma_j \leq 1 \text{ for all } (i, j) \in E\}.$$ 
Under this setup, the set of feasible schedules $S = \mathcal{I}(G)$.

1.3. Scheduling algorithm, performance metric. We need an algorithm to select schedule $\sigma(\tau) \in S$ for all $\tau \in \mathbb{N}$. From the perspective of network performance, we would like the scheduling algorithm such that the queues in network remain as small as possible given the arrival process.

Now, we formalize the notion of performance metric. In the setup described above, we define capacity region $\mathcal{C} \subset [0, 1]^n$ as the convex hull of the feasible scheduling set $S$, i.e.
$$\mathcal{C} = \left\{ \sum_{\sigma \in S} \alpha_\sigma \sigma : \sum_{\sigma \in S} \alpha_\sigma = 1, \text{ and } \alpha_\sigma \geq 0 \text{ for all } \sigma \in S \right\}.$$ 
The intuition behind this definition of capacity region comes from the fact that any algorithm has to choose schedule from $\mathcal{I}(G)$ each time and hence the time average of the ‘service rate’ induced by any algorithm must belong to $\mathcal{C}$. Therefore, if arrival rates $\lambda$ can be ‘served’ by any algorithm then it must belong to $\mathcal{C}$. Motivated by this, we call an arrival rate vector $\lambda$ admissible if $\lambda \in \Lambda$, where
$$\Lambda = \left\{ \lambda \in \mathbb{R}_+^n : \lambda \leq \sigma \text{ componentwise, for some } \sigma \in \mathcal{C} \right\}.$$ 
We say that an arrival rate vector $\lambda$ is strictly admissible if $\lambda \in \Lambda^o$, where $\Lambda^o$ is the interior of $\Lambda$ formally defined as
$$\Lambda^o = \left\{ \lambda \in \mathbb{R}_+^n : \lambda < \sigma \text{ componentwise, for some } \sigma \in \mathcal{C} \right\}.$$ 
Equivalently, we may say that the network is underloaded. Now we are ready to define the performance metric for a scheduling algorithm.

Definition 1.1 (Throughput optimal) We call a scheduling algorithm throughput optimal or provide 100% throughput or stable, if for any $\lambda \in \Lambda^o$ the underlying network Markov process (or queue-size vector) is Positive Recurrent\(^1\).

\(^1\)Algorithm itself may not be Markovian and in that case, queue-size vector may not be Markovian. In such situation, interpret throughput optimality as algorithm under which there is a well-defined notion of “limiting distribution”.

imsart-aap ver. 2007/04/13 file: lowerbound.tex date: September 22, 2008
The Throughput optimality is the first-order performance criteria. In essence it suggests that eventually queue-sizes remain finite. However, this ‘finite’ queue-sizes can be arbitrarily large. For this reason, we consider the second-order performance metric. To this end, restrict attention to scheduling algorithms that are Throughput optimal and let there be unique stationary distribution of queue-sizes under such algorithm. Let $\mathbb{E}[\sum Q_i]$ denote the net average queue-size with respect to this stationary distribution.

**Definition 1.2 (Queue-size optimal)** We call a scheduling algorithm queue-size optimal if for any $\lambda \in \Lambda^o$ the stationary distribution of queue-size vector exists and the net average queue-size with respect to this stationary distribution is the minimal.

1.4. Question of interest and contribution. In general, our interest is in finding queue-size optimal algorithms for general network model described above. Clearly, solving this problem exactly seems far from the reach of current understanding. A plausible approach to solve this problem requires the following: (a) lower bound on the average queue-size for any algorithm; and (b) an algorithm that has queue-sizes that are close to this lower bound. In the literature, to the best of our knowledge, such lower bounds are not available. For this reason, we are motivated to develop a method to obtain good lower bounds.

As the main result of this paper, we develop a new class of lower bound on the net average queue-size for any scheduling algorithm for the switched network model described above. This is summarized in Theorem 2.1.

To establish the non-triviality of our lower bound, we consider a class of wireless networks where the underlying graph has a specific structure (including bipartite graph, which is a model for wireless mesh network in a geographic area). For this setup, we show that the popular maximum weight algorithm (described next) has average queue-sizes that are within $O(1)$ of the lower bound (see Theorem 3.2). Thus, establishing constant factor optimality of the standard maximum weight scheduling algorithm for a non-trivial class of problems.

1.4.1. A popular algorithm. An important class of scheduling algorithms, known maximum-weight scheduling algorithm, was first proposed by Tassiulas and Ephremides [2]. These are myopic algorithms that utilize only the network state (queue-size) to decide the schedule. We describe the slotted time version of this algorithm. To this end, the algorithm changes decision in the beginning of every time slot using the $Q(\tau) = Q(\tau^+)$. Specifically,

$$\sigma(\tau) \in \arg \max_{\sigma \in S} \sum_i \sigma_i Q_i(\tau).$$
A generalized version of the MW algorithm, denoted by MW-f, picks a schedule

$$\sigma(\tau) \in \arg \max_{\sigma \in S} \sum_i \sigma_i f(Q_i(\tau)),$$

for some function f. It is well-known that MW-f algorithm has good throughput property as long as f : \( \mathbb{R}_+ \to \mathbb{R}_+ \) satisfies: (a) \( f(0) = 0 \), (b) f in strictly increasing and (c) \( f(rx) = rf(x) \) for any \( r, x > 0 \). To learn details, for example, see Stolyar [?] or Shah and Wischik [3].

2. Lower bound. We state our main result of lower bound on the average queue-size in this section. First, we define necessary definitions followed by the statement of the main result.

2.1. Workloads and virtual resources. Here we wish to identify the ‘virtual resources’ of a switched network that arise due to the scheduling constraint. As we shall see, the net queue-size (or workload) of these virtual resources will play an important role in identifying non-trivial lower bounds on the net queue-size of the system. To this end, for any \( \lambda \in \mathbb{R}_+^n \) consider an optimization problem PRIMAL(\( \lambda \)):

\[
\text{minimize} \quad \sum_{\pi \in S} \alpha_{\pi}
\]

over \( \alpha_{\pi} \in \mathbb{R}_+ \) for all \( \pi \in S \)

such that \( \lambda \leq \sum_{\pi \in S} \alpha_{\pi} \pi \) componentwise

This problem asks whether it is possible to find a combination of schedules which can serve arrival rates \( \lambda \); clearly \( \lambda \) is admissible if and only if the solution to the primal is \( \leq 1 \). Now consider its dual problem DUAL(\( \lambda \)):

\[
\text{maximize} \quad \xi \cdot \lambda
\]

over \( \xi \in \mathbb{R}_+^n \)

such that \( \max_{\pi \in S} \xi \cdot \pi \leq 1 \)

The solution is clearly attained when the constraint is tight. Given a queue size vector \( Q \) and any dual-feasible \( \xi \) satisfying the constraint with equality, call \( \xi \cdot Q \) the workload at the virtual resource \( \xi \). The virtual resource specifies a combination of several actual resources (namely the queues themselves). The long-run rate at which
work arrives at the virtual resource is \( \xi \cdot \lambda \), and the maximum rate at which it can be served is 1. Define \( \lambda_{\xi} = \xi \cdot \lambda \). If \( \lambda_{\xi} > 1 \), then the workload must increase to infinity, and if \( \lambda_{\xi} \leq 1 \) then there is a chance that the workload might be stable. A good scheduling algorithm tries to minimize these workloads simultaneously. However, any algorithm can do only ‘so much’ in terms of minimizing the workloads. The quantification of ‘so much’ and its effect on the overall system will lead to non-trivial lower bound on the net queue-size as stated in Theorem 2.1.

2.2. Main result. We state our main result on the lower bound of average queue-sizes for any scheduling algorithm that is stable with unique stationary distribution. This bound is applicable when the system is sufficiently “heavily” loaded.

**Theorem 2.1** Given a switched network, let \( Z = \{ \xi \in \{0, 1\}^n : \max_{\pi \in S} \xi \cdot \pi \leq 1 \} \). Then, for any scheduling algorithm that is stable and has unique stationary distribution of the queue-size \( Q(\cdot) \), the following holds with interpretation of \( \delta > 0 \) as a universal constant.

(a) For any \( \xi \in Z \) with \( \lambda_{\xi} \in (1 - \delta, 1] \), there exists a universal constant \( \Gamma > 0 \) such that with respect to the stationary distribution

\[
\mathbb{E}[Q \cdot \xi] \geq \frac{\Gamma}{1 - \lambda_{\xi}}.
\]

(b) More generally, the net average queue-size with respect to the stationary distribution is lower bounded by the following optimization problem:

\[
\begin{align*}
\text{minimize} & \quad q \cdot 1 \\
\text{over} & \quad q \in \mathbb{R}_+^n \\
\text{such that} & \quad q \cdot \xi \geq \frac{\Gamma}{1 - \lambda_{\xi}}, \quad \text{for all } \xi \in Z \text{ with } \lambda_{\xi} \geq 1 - \delta.
\end{align*}
\]

**Proof.** We state the proof of part (a). The proof of part (b) follows from part (a) and straightforward arguments. The following lemma will serve as a key to the result (a).

**Lemma 2.2** Given \( \alpha \geq 1 \), there exists a constant \( t(\alpha) > 0 \) and \( \beta \in (0, 1) \) such that for any Poisson process of rate \( \lambda \in [1/2, 1] \) and \( t \geq t(\alpha) \),

\[
\Pr \left[ \frac{A_\lambda(t) - \lambda t}{\sqrt{\lambda t}} \geq \alpha \right] \geq \beta.
\]

Here \( A_\lambda(t) \) denote the cumulative arrivals in interval \([0, t]\) under Poisson process of rate \( \lambda \).
Proof. Consider a fixed $\lambda \in [1/2, 1]$, say $\lambda = 1/2$. Define

\[
 f_{1/2}(\alpha, t) = \Pr \left[ \frac{A_{1/2}(t) - t/2}{\sqrt{t/2}} \geq \alpha \right].
\]

By Central Limit Theorem style estimate, it follows that

\[
 \lim_{t \to \infty} f_{1/2}(\alpha, t) \to 2\beta,
\]

for some constant $\beta > 0$. That is, there exists $\hat{t}(\alpha)$ such that for $t \geq \hat{t}(\alpha)$ we have $f_{1/2}(\alpha, t) \geq \beta$. Now by property of Poisson process it follows that $A_{\lambda}(t)$ is distributionally same as $A_{1/2}(2\lambda t)$ for all $t > 0$. Therefore, it follows that for $t \geq \hat{t}(\alpha)/2\lambda$, $A_{\lambda}(t)$ is equal (in distribution) to $A_{1/2}(2\lambda t) = A_{1/2}(s)$ with $s = 2\lambda t \geq \hat{t}(\alpha)$. Therefore, from above discussion we obtain that for $t \geq \hat{t}(\alpha)/2\lambda$,

\[
 \Pr \left[ \frac{A_{\lambda}(t) - \lambda t}{\sqrt{\lambda t}} \geq \alpha \right] = \Pr \left[ \frac{A_{1/2}(s) - s/2}{\sqrt{s/2}} \geq \alpha \right],
\]

with $s \geq \hat{t}(\alpha)$. Therefore, it is larger than $\beta$. The proof is completed by selecting $\hat{t}(\alpha) = \hat{t}(\alpha)$.

Now we are ready to prove Theorem 2.1(a). To this end, note that we are interested in scheduling algorithms under which vector of queues $Q(\cdot)$ has well-defined stationary distribution. Consider this system being stationary at time 0 (assume system started at time $-\infty$ for ease of this part of the proof contrary to our setup where we assumed system starts at time 0). Now, $Q(0)$ must satisfy the following standard queuing equation: for any $\tau \in \mathbb{N}$,

\[
 Q(0) \geq \left( A(-\tau, 0) - \sum_{s=-\tau}^{1} \sigma(s) \right)^{+},
\]

for some collection of schedule $\sigma(s) \in \mathcal{S}$ for $s \in [-\tau, 0)$; $A(-\tau, 0)$ denotes the arrivals in time interval $(-\tau, 0)$. Therefore, for any $\tau \in \mathbb{N}$ and $\xi \in \mathcal{Z}$,

\[
 \mathbb{E}[Q(0) \cdot \xi] \geq \mathbb{E} \left[ \left( A(-\tau, 0) \cdot \xi - \sum_{s=-\tau}^{1} \sigma(s) \cdot \xi \right)^{+} \right],
\]

Now, for $\xi \in \mathcal{Z}$, by definition $\max_{\pi \in \mathcal{S}} \xi \cdot \pi \leq 1$. Therefore, it follows that

\[
 \mathbb{E}[Q(0) \cdot \xi] \geq \mathbb{E} \left[ (A(-\tau, 0) \cdot \xi - \tau)^{+} \right].
\]

Now, we shall consider $\mathbb{E} \left[ (A(-\tau, 0) \cdot \xi - \tau)^{+} \right]$ for particular choice for $\tau$. To this end, consider $A(-\tau, 0) \cdot \xi$. Since $\xi \in \mathcal{Z}$ and hence $\xi \in \{0, 1\}^{n}$, $A(-\tau, 0) \cdot \xi$ corresponds to
adding arrival processes to a subset of queues. Since arrival process to each queue is independent Poisson process and by property of Poisson process that merging of independent Poisson processes results into another Poisson process with rates being added up, $A(-\tau, 0) \cdot \xi$ corresponds to cumulative arrivals of a Poisson process of rate $\lambda_\xi = \lambda \xi$ in time interval of length $\tau$. By Lemma 2.2 it follows for that if $\lambda_\xi \in [1/2, 1]$ then for any $\tau \geq t(\alpha)$ we have

$$\Pr \left( A(-\tau, 0) \cdot \xi \geq \lambda_\xi \tau + \alpha \sqrt{\lambda_\xi \tau} \right) \geq \beta > 0.$$  

Putting all the above discussion together, it follows that for $\tau \geq t(\alpha)$, we have

$$E \left[ (A(-\tau, 0) \cdot \xi - \tau)^+ \right] \geq \beta \left( \lambda_\xi - 1 \right) \tau + \alpha \sqrt{\lambda_\xi \tau}.$$

Select $\tau^* = \frac{\alpha^2 \lambda_\xi}{8(1 - \lambda_\xi)^2}$. Clearly, there exists $\delta > 0$ so that for any $\lambda_\xi \in (1 - \delta, 1]$, thus selected $\tau^* > t(\alpha)$. For such a selection of $\tau = \tau^*$ in (1) it follows that

$$E \left[ (A(-\tau, 0) \cdot \xi - \tau)^+ \right] \geq \left( \frac{\sqrt{8} - 1}{8} \right) \frac{\beta \alpha^2 \lambda_\xi}{1 - \lambda_\xi}.$$

From (2) it follows that there is a universal constant $\Gamma$ so that the claimed lower bound in Theorem 2.1(a) is satisfied. This completes the proof of Theorem 2.1(a).

3. Non-triviality of lower bound: an example. We apply the lower bound of Theorem 2.1 to the example of wireless network model with independent set constraints to establish its non-triviality. As defined earlier, let $G = (V, E)$ be the network graph. Let $\chi(G)$ denote the chromatic number of $G$, i.e. minimum number of colors required to color all vertices $V$ of $G$ so that no two neighbors have the same color; let $\alpha(G)$ denote the independence number of $G$, i.e. the size of the large independent set in $G$. Clearly, $\chi(G) \alpha(G) \geq n$ since vertices colored by the same color form an independent set. We state the following result about the standard maximum weight algorithm, which follows from the well-known moment bounds based on Lyapunov function.

**Proposition 3.1** Consider a wireless network $G$ with $\lambda$ such that $\rho(\lambda) = \text{PRIMAL}(\lambda) < 1$. Then, under the MW algorithm the queue-size vector has unique stationary distribution such that

$$E[Q \cdot 1] \leq \frac{\chi(G)}{2(1 - \rho(\lambda))} [2\lambda \cdot 1 + \alpha(G)].$$

**Proof.** The standard proof of stability for MW algorithm uses the quadratic Lyapunov function, $L(\tau) = L(Q(\tau)) = \sum_i Q_i^2(\tau)$. Specifically, it can be shown (after some algebraic manipulation) that

$$E[L(\tau + 1) - L(\tau) | Q(\tau)] \leq 2(Q(\tau) \lambda - Q(\tau) \cdot \sigma(\tau)) + 2\lambda \cdot 1 + \alpha(G).$$
The above uses facts: (a) for Poisson process, $\mathbb{E}[A^2_i(1)] = \lambda_i^2 + \lambda_i$, and (b) $\sigma(\tau) \cdot \mathbf{1} \leq \alpha(G)$ for all $\tau$. Now it follows by definition of $\rho(\lambda)$ and fact that $Q(\tau) \cdot \sigma(\tau)$ is maximum among all independent sets, we have that

\[ (Q(\tau) \cdot \lambda - Q(\tau) \cdot \sigma(\tau)) \leq -(1 - \rho(\lambda))Q(\tau) \cdot \sigma(\tau). \]

Now since chromatic number is $\chi(G)$, there exists $\chi(G)$ independent sets that cover all $n$ queues. Therefore, it follows that

\[ Q(\tau) \cdot \sigma(\tau) \geq \frac{1}{\chi(G)} Q(\tau) \cdot \mathbf{1}. \]

Putting all together, we obtain that for all $\tau$,

\[ \mathbb{E}[L(\tau + 1) - L(\tau)|Q(\tau)] \leq \frac{2(1 - \rho(\lambda))}{\chi(G)} Q(\tau) \cdot \mathbf{1} + 2\lambda \cdot \mathbf{1} + \alpha(G). \]

Using standard Lyapunov-Foster’s method for moment bound it follows that

\[ \mathbb{E}[Q \cdot \mathbf{1}] \leq \frac{\chi(G)}{2(1 - \rho(\lambda))} [2\lambda \cdot \mathbf{1} + \alpha(G)]. \]

This completes the proof of Proposition 3.1.

The following result establishes the optimality of MW-algorithm and non-triviality of lower bound of Theorem 2.1 for a (non-trivial) class of graphs.

**Theorem 3.2** Consider $G = (V, E)$ such that it is bipartite and there exists subgraph $G' = (V, E')$ with $E' \subset E$ so that $G'$ is $d$-regular for some $d \geq 1$. Then, there exists sequence of $\lambda$ with $\rho(\lambda)$ taking any value in $(1 - \delta, 1]$ such that for any algorithm with stationary distribution of queue-size,

\[ \mathbb{E}[Q \cdot \mathbf{1}] \geq \frac{\Gamma n}{2(1 - \rho(\lambda))}. \]

And, under the MW algorithm for any $\lambda$ with $\rho(\lambda) < 1$ we have

\[ \mathbb{E}[Q \cdot \mathbf{1}] \leq \frac{3n}{1 - \rho(\lambda)}. \]

Thus, Theorem 3.2 implies that the MW algorithm is $3/2\Gamma = O(1)$ optimal in terms of average queue-size.

**Proof.** The proof of Theorem 3.2 follows through an application of Theorem 2.1 and Proposition 3.1. First, the lower bound. For this consider arrival rate vector $\lambda = \frac{\rho(\lambda)}{2} \mathbf{1}$ with $\rho(\lambda) \in (1 - \delta, 1)$. Then, for any $(i, j) \in E$,

\[ \lambda_i + \lambda_j = \rho(\lambda)/2 + \rho(\lambda)/2 = \rho(\lambda) < 1. \]
Now for bipartite graph $G$, it is well known that the convex hull of independent sets is equivalent to set

$$S(G) = \{ \mathbf{x} \in [0,1]^n : x_i + x_j \leq 1, \ \forall \ (i,j) \in E \}.$$ 

Therefore, $\lambda$ as defined above is indeed strictly admissible.

Now, we consider a special collection of "virtual resources" that will lead to a desired lower bound. To this end, for each $e' \in E'$, consider $\xi_{e'} \in \{0,1\}^n$ which has exactly two co-ordinate 1 and rest 0; its $i'$ and $j'$ co-ordinates are 1 where $e' = (i', j') \in E'$. For each such $\xi_{e'}$ it follows that $\lambda \xi_{e'} = \rho(\lambda)$ and $Q \xi_{e'} = Q_{i'} + Q_{j'}$. Therefore, Theorem 2.1 implies that for each $e' \in E'$,

$$\mathbb{E}[Q \cdot \xi_{e'}] \geq \frac{\Gamma}{1 - \rho(\lambda)}.$$ 

Summing over all $e' \in E'$ and recalling the fact that $G'$ is a $d$-regular graph (i.e. under $E'$ degree of each node is $d$), we have that

$$d \mathbb{E}[Q \cdot 1] \geq \frac{\Gamma |E'|}{1 - \rho(\lambda)}.$$ 

But, $2|E'| = dn$. Therefore, we obtain the desired lower bound:

$$\mathbb{E}[Q \cdot 1] \geq \frac{\Gamma n}{2(1 - \rho(\lambda))}.$$ 

To obtain the upper bound, note that for a bipartite graph $G$, $\chi(G) = 2$; due to independent set constraint $\lambda \cdot 1 \leq n$ and $\alpha(G) \leq n$. Using these values in Proposition 3.1, the upper bound follows.

4. Discussion. This note presented a new method for lower bounding the average queue-sizes for a general model of switched network. The non-triviality of lower bound was established by means of an example network – the wireless network. Specifically, when the network graph has certain properties (bipartite with a regular subgraph), our lower bound establishes the constant factor optimality of the standard MW algorithm. This example is only a tip of an ice-berg we believe that the lower bounding technique introduced here will have much wider impact in the analysis and design of scheduling algorithms for switched networks in general.

References.


