Three metrics for stochastic networks: capacity, queue-size and complexity

Devavrat Shah
Email: devavrat@mit.edu
LIDS and ORC
Dept. of EECS, MIT

***Abstract*—We consider a class of constrained queuing network called switched networks, important sub-class of the stochastic processing network (cf. Harrison (2000)), in which there are constraints on which queues can be served simultaneously. Such networks have served as effective models for understanding various types of dynamic resource allocation problems arising in communication networks like the Internet, computer architecture, manufacturing, etc. In such systems, a scheduling policy is required to make resource allocation decisions in terms of which queues to serve at each time subject to constraints. The performance of the system is crucially determined by the scheduling policy. The performance of such scheduling policies, measured with respect to the following three metrics is of utmost importance: (a) capacity, (b) induced queue-sizes or latency, and (c) complexity of the implementation of the policy. Ideally, one is interested in determining the trade-offs achievable between these metrics characterized through the *pareto* performance boundary. In this note, we shall summarize the state-of-art along with recent progresses towards this somewhat ambitious program.**

***Index Terms*—switched network, input-queued switch, scheduling, maximum weight scheduling, fluid models, state space collapse, heavy traffic, diffusion approximation**

I. Model, A class of Policies

We consider a collection of N queues operating in discrete time, indexed by $\tau \in \{0, 1, \dots\}$: $Q_n(\tau)$ be work in queue n, $1 \leq n \leq N$ at time τ, $\mathbf{Q}(\tau) = [Q_n(\tau)]$ be the vector of queue-sizes and initially it is $\mathbf{Q}(0)$. Let $A_n(\tau)$ be the total amount of work arriving to queue n and $B_n(\tau)$ be the cumulative potential service provided to queue n, up to time τ respectively, with $\mathbf{A}(0) = \mathbf{B}(0) = \mathbf{0} = [0]$. We consider single-hop network (for simplicity of exposition). Let $d\mathbf{A}(\tau) = \mathbf{A}(\tau+1) - \mathbf{A}(\tau)$ and $d\mathbf{B}(\tau) = \mathbf{B}(\tau+1) - \mathbf{B}(\tau)$. Then basic Lindley recursion is

$$\mathbf{Q}(\tau+1) = \big[\mathbf{Q}(\tau) - d\mathbf{B}(\tau)\big]^+ + d\mathbf{A}(\tau) \tag{1}$$

where the $[\cdot]^+$ is taken componentwise. The fundamental 'switched network' constraint is that there is some finite set $\mathcal{S} \subset \mathbb{R}_+^N$ such that

$$d\mathbf{B}(\tau) \in \mathcal{S} \quad \text{for all } \tau. \tag{2}$$

For simplicity, we shall consider $\mathcal{S} \subset \{0,1\}^N$. We will refer to $\boldsymbol{\pi} \in \mathcal{S}$ as a schedule, and $\mathcal{S}$ as the set of allowed schedules. The departure from queue n up to time τ is

$$D_n(\tau) = \sum_{s=0}^{\tau} dB_n(\tau)\mathbf{1}_{\{Q_n(\tau)>0\}}, \tag{3}$$

where $\mathbf{1}_{\{x\}} = 1$ if $x =$ true and 0 otherwise.

978-1-4244-8953-4/11/$26.00

A policy needs to choose schedule $d\mathbf{B}(\tau) \in \mathcal{S}$ in each time slot. The specific class of policies of interest are the so called maximum weight (MW) introduced (in basic version) by Tassiulas and Ephremides [10]. In the basic version, the schedule is chosen as follows (ties broken randomly or as per a fixed rule):

$$d\mathbf{B}(\tau) \in \operatorname*{argmax}_{\boldsymbol{\pi} \in \mathcal{S}} \boldsymbol{\pi} \cdot \mathbf{Q}(\tau), \tag{4}$$

where we use notation: $\mathbf{u} \cdot \mathbf{v} = \sum_n u_n v_n$ for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$. More generally, given an increasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = 0$, $f(x) \to \infty$ as $x \to \infty$, the MW-f policy chooses schedule

$$d\mathbf{B}(\tau) \in \operatorname*{argmax}_{\boldsymbol{\pi} \in \mathcal{S}} \boldsymbol{\pi} \cdot f(\mathbf{Q}(\tau)). \tag{5}$$

The specific class of policies of interest are those induced by $f(x) = x^\alpha$ for $\alpha \in (0, 1)$; we shall denote this policy as MW-α policy. We shall assume the following.

Assumption 1.1: We assume that $\mathcal{S} \subset \{0,1\}^N$ is monotone: if $\boldsymbol{\pi} \in \mathcal{S}$, then for any $\boldsymbol{\rho} \in \{0,1\}^N$ with $\rho_n \in \{0, \pi_n\}$, $\boldsymbol{\rho} \in \mathcal{S}$.

To evaluate the performance of the system, we model the uncertainty in the system (in form of arrivals) by means of appropriate stochastic model. Specifically, we assume the following.

Assumption 1.2: Let $\mathbf{A}(\cdot)$ be a random process with stationary increments. Let there be $\boldsymbol{\lambda} \in \mathbb{R}_+^N$ so that for any $r \in \mathbb{Z}_+$

$$\mathbb{P}\Big(\sup_{\tau \leq r} \frac{1}{r}\Big|\mathbf{A}(\tau) - \boldsymbol{\lambda}\tau\Big| \geq \varepsilon_r\Big) \leq \delta_r, \tag{6}$$

where $\varepsilon_r,\ \delta_r$ go to 0 as $r \to \infty$.

The specific instances of $\mathbf{A}(\cdot)$ that satisfy Assumption 1.2 and we shall consider are: (a) $A_n(\cdot)$ is Bernoulli process with parameter $\lambda_n \in [0, 1]$ independent across n; (b) $A_n(\cdot)$ is Poisson process with parameter $\lambda_n \in \mathbb{R}_+$ independent across n.

II. Results

We describe the known results about interplay between three performance metrics of scheduling policies: (a) capacity, (b) average queue-size as well as exponential tail bounds on queue-sizes, and (c) complexity of implementation of the policy.

A. Capacity.

This is about how well the network resource is utilized by the scheduling policy. While there are various ways to understand this, we shall define the effective resource utilization by studying the net departure rate. Specifically, define

$$d_n = \liminf_{\tau} \frac{1}{\tau} D_n(\tau). \tag{7}$$

The net departure rate is $\bar{d} = \sum_n d_n$. An algorithm is called optimal in terms of capacity if for any given $\boldsymbol{\lambda} \in \mathbb{R}_+^N$, the induced net departure rate $\bar{d} = \bar{d}(\boldsymbol{\lambda})$ is maximal possible with probability 1. In [9], the following is established (by means of fluid model):

1. For any policy with probability 1, $\bar{d}(\boldsymbol{\lambda}) \leq \big(\sum_n \lambda_n\big) - \bar{q}(\boldsymbol{\lambda})$ where $\bar{q}(\boldsymbol{\lambda})$ is the value of the following optimization problem:

$$\begin{aligned} &\text{minimize} \sum_n r_n \quad \text{over } \mathbf{r} \in \mathbb{R}_+^N \\ &\text{subject to } \zeta \cdot \mathbf{r} \ \geq \zeta \cdot \boldsymbol{\lambda} - 1, \ \ \forall \ \ \zeta \in \mathcal{D}, \end{aligned} \tag{8}$$

   where $\mathcal{D} = \{\zeta \in \mathbb{R}_+^N : \zeta \cdot \boldsymbol{\pi} \leq 1 \ \forall \ \boldsymbol{\pi} \in \mathcal{S}\}$.
2. The net departure rate induced by MW-α is, with probability 1, $\bar{d}_\alpha(\boldsymbol{\lambda}) = \big(\sum_n \lambda_n\big) - \bar{q}_\alpha(\boldsymbol{\lambda})$ where $\bar{q}_\alpha(\boldsymbol{\lambda})$ is the value of the following optimization problem:

$$\begin{aligned} &\text{minimize} \sum_n r_n^{1+\alpha} \quad \text{over } \mathbf{r} \in \mathbb{R}_+^N \\ &\text{subject to } \zeta \cdot \mathbf{r} \ \geq \zeta \cdot \boldsymbol{\lambda} - 1, \ \ \forall \ \ \zeta \in \mathcal{D}, \end{aligned} \tag{9}$$

   where $\mathcal{D} = \{\zeta \in \mathbb{R}_+^N : \zeta \cdot \boldsymbol{\pi} \leq 1 \ \forall \ \boldsymbol{\pi} \in \mathcal{S}\}$.
3. Therefore, for any $\boldsymbol{\lambda} \in \mathbb{R}_+^N$,

$$\lim_{\alpha \downarrow 0} \bar{q}_\alpha(\boldsymbol{\lambda}) = \big(\sum_n \lambda_n\big) - \bar{q}(\boldsymbol{\lambda}).$$

The above results implies that the MW-α policy is asymptotically optimal in terms of maximizing the departure rate as $\alpha \to 0^+$.

B. Queue-size on average.

Here interest is in understanding the behavior of average queue-size induced by policy when the notion of stationary distribution as well as average queue-size (with respect to it) are well defined. To state results (in a clean form) about average queue-sizes, we shall assume arrival process to be independent Poisson. Given rate vector $\boldsymbol{\lambda} \in \mathbb{R}_+^N$, define the load $\rho(\boldsymbol{\lambda})$ as

$$\begin{aligned} &\text{minimize} \sum_{\boldsymbol{\pi} \in \mathcal{S}} \alpha_{\boldsymbol{\pi}} \text{over} \quad \alpha_{\boldsymbol{\pi}} \geq 0, \ \forall \ \boldsymbol{\pi} \in \mathcal{S} \\ &\text{subject to} \quad \boldsymbol{\lambda} \ \leq \sum_{\boldsymbol{\pi}} \alpha_{\boldsymbol{\pi}} \boldsymbol{\pi}. \end{aligned} \tag{10}$$

Clearly, $\mathbf{Q}(\cdot)$ is a Markov chain under MW-α policy when arrival process is Poisson. When $\rho(\boldsymbol{\lambda}) < 1$ then it is positive recurrent with well defined, unique stationary distribution and $\rho(\boldsymbol{\lambda}) < 1$ is necessary for this. For any MW-α policy with $\alpha > 0$, $\mathbb{E}[\sum_n Q_n]$ with respect to this stationary distribution is well defined (see [6]). Using the Foster-Lyapunov moment bound, it follows that for any system considered here, under the MW-1 policy, the stationary average queue-size is bounded above as

$$\mathbb{E}\Big[\sum_n Q_n\Big] \leq C \frac{n^2}{1 - \rho(\boldsymbol{\lambda})}, \tag{11}$$

where $C > 0$ is a universal constant. In [8] the following $\mathcal{S}$ dependent lower bound on the net average queue-size under any policy is established:

$$\begin{aligned} &\text{minimize} \sum_n r_n \quad \text{over } \mathbf{r} \in \mathbb{R}_+^N \\ &\text{subject to } \zeta \cdot \mathbf{r} \ \geq \frac{\boldsymbol{\lambda} \cdot \zeta^2}{2(1 - \boldsymbol{\lambda} \cdot \zeta)}, \ \ \forall \ \ \zeta \in \mathcal{D}, \end{aligned} \tag{12}$$

where $\mathcal{D} = \{\zeta \in \mathbb{R}_+^N : \zeta \cdot \boldsymbol{\pi} \leq 1 \ \forall \ \boldsymbol{\pi} \in \mathcal{S}\}$.

The overall challenge is to identify $\chi_1(\mathcal{S})$ that depends on $\mathcal{S}$ so that under the best policy with well defined stationary distribution for any $\rho \in (0, 1)$,

$$\sup_{\boldsymbol{\lambda} : \rho(\boldsymbol{\lambda}) = \rho} \mathbb{E}\big[\sum_n Q_n\big] = \Theta\Big(\frac{\chi_1(\mathcal{S})}{1 - \rho}\Big). \tag{13}$$

Clearly, the above statement implicitly conjectures existences of such $\chi_1(\mathcal{S})$ that determines the optimal (up to universal constants) dependence of average queue-size on scheduling constraints $\mathcal{S}$ for any such system. The (11) and (12) provide upper and lower bound on such quantity. In [8] it is shown that the $\chi_1(\mathcal{S})$ (up to constants) is characterized by the lower bound (12) for the (general enough) instance of switched network model induced by wireless network model (see Section II-D for description of this model) with *regular enough* constraint graph structure describing $\mathcal{S}$. Indeed, we believe that bound of (12) is reasonably accurate.

C. Exponential tail probability.

Here interest is in understanding the further detailed behavior of the stationary distribution of queue-size when it exists. Specifically, we shall assume that arrival process is Bernoulli with rate vector $\boldsymbol{\lambda} \in [0, 1]^N$ so that $\rho(\boldsymbol{\lambda}) < 1$. Again, in this setup the $\mathbf{Q}(\cdot)$ forms a Markov chain under MW-α policy for any $\alpha > 0$. It is positive recurrent with unique stationary distribution as long as $\rho(\boldsymbol{\lambda}) < 1$. In [6], it is shown that with respect to this stationary distribution, under MW-α policy

$$\limsup_{x \to \infty} \frac{1}{x} \log \mathbb{P}\Big(\|\mathbf{Q}\|_{1+\alpha} \geq x\Big) \leq -(1 - \rho(\boldsymbol{\lambda})) N^{-\frac{2+\alpha}{1+\alpha}}. \tag{14}$$

Further, for any policy it can be shown that for each $\rho \in (0, 1)$ there exists $\boldsymbol{\lambda}$ such that $\rho(\boldsymbol{\lambda}) = \rho$ and

$$\liminf_{x \to \infty} \frac{1}{x} \log \mathbb{P}\Big(\|\mathbf{Q}\|_{1+\alpha} \geq x\Big) \geq -(1 - \rho) N^{\frac{\alpha}{1+\alpha}}. \tag{15}$$

Similar to the average queue-size, the overall challenge is to identify $\chi_2(\mathcal{S}, \alpha)$ that depends on $\mathcal{S}$ (and $\alpha > 0$) so that under the best policy with well defined stationary distribution for any $\rho \in (0, 1)$,

$$\begin{aligned} \sup_{\boldsymbol{\lambda} : \rho(\boldsymbol{\lambda}) = \rho} \ \liminf_{x \to \infty} \frac{1}{x} \log \mathbb{P}\Big(\|\mathbf{Q}\|_{1+\alpha} \geq x\Big) &\geq -C_1 (1 - \rho) \chi_2(\mathcal{S}, \alpha) \\ \sup_{\boldsymbol{\lambda} : \rho(\boldsymbol{\lambda}) = \rho} \ \limsup_{x \to \infty} \frac{1}{x} \log \mathbb{P}\Big(\|\mathbf{Q}\|_{1+\alpha} \geq x\Big) &\leq -C_2 (1 - \rho) \chi_2(\mathcal{S}, \alpha), \end{aligned} \tag{16}$$

where $0 < C_1 \leq C_2$ are some universal constants. Clearly, the above statement implicitly conjectures existences of such $\chi_2(\mathcal{S}, \alpha)$ that determines the optimal (up to universal constants) dependence of exponential tail bound of $1+\alpha$ norm of queue-sizes with respect to the stationary distribution. Indeed, the (14) and (15) provide upper and lower bound on such quantity.

D. Complexity.

The summary of results thus far suggest that the MW-α class of policies are reasonably good: MW-0^+ policy is (near) optimal in terms of capacity, it has good scaling in terms of average queue-size and very good behavior in terms of exponential tail probability. And most importantly, it is a *myopic* policy, i.e. it uses only current network state (queue-sizes) to make the scheduling decision. Therefore, it is only natural to wonder: *do we have a reasonable answer for the problem of designing scheduling policies for switched networks?*

To answer this question, it is important to understand the context where such policies will be used. Specifically, in the application domain like communication networks, the implementation of policies is highly constrained: (i) it should be very simple in terms of computation and data structure requirement so that it can operate at very high aggregate bandwidth (say making decision once every few nano seconds !) with existing limitations on the memory bandwidth with limited hardware requirement at low power; (ii) it should be preferably iterative and distributed so as to allow for architectures that are scalable. Such stringent constraints immediately lead to the following questions: is it possible to implement the MW policy with above requirements for arbitrary scheduling set $\mathcal{S}$? if not, how does the performance suffer from implementation limitation ?

Before we start answering these questions rigorously, it is important to lay down constraints explicitly. The main problem is that the precise constraints are problem dependent and the above stated constraints, while provide flavor of the problem, do not define them explicitly. For that reason, collectively over the past decade or so, researchers have focused on few concrete applications where the constraints are quite clear. One such important problem has been the design of medium access in wireless networks. Formally, there are N wireless transmitters or nodes or queues, denoted by $V = \{1, \dots, N\}$. Let $E \subset V \times V$ represent the scheduling or transmission constraints: if $(i, j) \in E$ then no two nodes/queues can transmit/be served simultaneously. Thus, the space of all possible schedules is

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

The implementation of any policy for making transmission decision is highly constrained: (C1) each node must make decision (using few, preferably constant number of computations) to transmit or not in each time slot on its own based on local observations/historical information possibly summarized through limited data structure; (C2) nodes have (so called delayed carrier sense) information about transmission of other nodes in prior time slots (and hence if they transmitted, they know whether their transmissions were successful or not). Indeed, the popular implemented protocols that go by the names *Aloha*, *Random Backoff*, etc. satisfy C1 and C2.

Now in this setup, the maximum weight policy requires solving the so called maximum weight independent set problem in graph G (with weights dependent on queue-size). In general, this is hard problem and hence it is not clear if one can implement the maximum weight policy as is with stringent constraints like C1 and C2. This led to a long line of research finally resulting in its resolution very recently by Shah and Shin [3], [4] and Jiang and Walrand [2], [1]: they provide implementation so that the resulting policy keeps network Markov process positive recurrent as long as the system is underloaded (i.e. $\rho(\boldsymbol{\lambda}) < 1$). However, the performance of such implementation in terms of queue-size scaling is not clear.

As mentioned above, the MW-1 policy for example, induced average queue-size with respect to stationary distribution that scales as $O(N^2/(1-\rho(\boldsymbol{\lambda}))$ when $\rho(\boldsymbol{\lambda}) < 1$. And the best bound one can obtain on average queue-size under the above mentioned implementations are scaling super polynomially in N, at the best.

In summary, MW is capacity achieving, provides small (polynomially scaling in N) average queue-size but may not yield to implementation that has (time-)computational complexity scaling polynomial in N (C1, C2 is too much to ask for !). The implementations of [3], [4], [2], [1] have polynomial complexity (actually, they satisfy C1, C2), is capacity achieving (positive recurrence when $\rho(\boldsymbol{\lambda}) < 1$) but may not yield small (polynomially scaling in N) average queue-sizes. This leads to the following fundamental question: *is it possible to have an implementation that is (i) capacity achieving, (ii) has small (polynomial in N) queue-sizes, and (iii) has polynomial in N computational complexity?*.

Recently, Shah, Tse and Tsitsiklis [5] established that the answer to this question is *NO* for arbitrary G (assuming the standard computational hypothesis). The precise statement of the result, details as well as extensions can be found in [5].

III. Conclusion

In conclusion, we would like to take note of the key results and open problems surveyed here. Specifically, we surveyed performance of scheduling policies for switched networks in the context of three metrics: capacity, queue-size and complexity. The MW-α policy as $\alpha \downarrow 0$ achieves the maximal capacity, defined in terms of the effective net departure rate for any switched network. This is a strong evidence of the qualitative conjecture about goodness of MW-0^+ policy in the MW class of algorithms originally introduced by Shah and Wischik [7]. The MW family of policies induce stationary average queue-sizes when system is underloaded: they are reasonably good and close to fundamental, policy independent, lower bounds on them; similar qualitative result holds for the exponential tail probability. Though MW policies are mypoic and have very good performance in terms of capacity and queue-size, in general their implementation can require solving computationally hard problems. Indeed, for general switched network, it is not possible to have any policy that is computationally simple and has good performance in terms of queue-size in complexity.

The questions of determining exact scaling of average queue-size (i.e. $\chi_1(\mathcal{S})$ and exponential tail probability (i.e. $\chi_2(\mathcal{S})$) remain important quest going forward – resolution of these will definitely advance the frontiers of methods for large complex queuing networks. And the grand challenge is to develop framework to understand the *pareto* boundary reflecting interplay between these three performance metrics.

REFERENCES

[1] JIANG, L., SHAH, D., SHIN, J., AND WALRAND, J. Distributed random access algorithm: Scheduling and congestion control.

[2] JIANG, L. AND WALRAND, J. (2008). A distributed csma algorithm for throughput and utility maximization in wireless networks. In *Proceedings of 46th Allerton Conference on Communication, Control, and Computing, Urbana-Champaign, IL.*

[3] RAJAGOPALAN, S., SHAH, D., AND SHIN, J. (2009). A network adiabatic theorem: an efficient randomized protocol for contention resolution. In *ACM Sigmetrics/Performance*.

[4] SHAH, D. AND SHIN, J. (2009). Randomized Scheduling Algorithm for Queueing Networks. *Arxiv preprint arXiv:0908.3670.*

[5] SHAH, D., TSE, D. N. C., AND TSITSIKLIS, J. N. (2009). Hardness of low delay scheduling. *Submitted to IEEE Transaction on Information Theory*.

[6] SHAH, D., TSITSIKLIS, J. N., AND ZHONG, Y. (2010). Qualitative properties of α-weighted policies. In *ACM Sigmetrics/Performance*.

[7] SHAH, D. AND WISCHIK, D. J. (2006, 2007). Switched networks with maximum weight policies: Fluid approximation and multiplicative state space collapse. *Submitted to Annals of Applied Probability*.

[8] SHAH, D. AND WISCHIK, D. J. (2008). Lower bound and optimality in switched networks. In *Allerton conference on communication, control and computing*.

[9] SHAH, D. AND WISCHIK, D. J. (2010). Fluid model of switched networks in overload. *Submitted to QUESTA*.

[10] TASSIULAS, L. AND EPHREMIDES, A. (1992). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control 37*, 1936–1948.