Utility Maximization for Arbitrary Traffic in Communication Networks

Jin-Ghoo Choi, Wooguil Pak, and Changhee Joo

***Abstract*—MaxWeight-based throughput-optimal link scheduling schemes has been shown to attain the maximum utility in communication networks for any feasible input traffic. However, for input traffic outside the capacity region, they fail to achieve the maximum utility. Although a joint solution with admission controller has been recently developed to address the problem, it requires additional structural complexity with non-intuitive state variables, which further complicates the already complex MaxWeight-based schemes. In this paper, we take more direct control of the arrivals, and develop a new joint solution with simpler admission controller and less structural complexity. Our scheme provably achieves the maximum utility, and empirically outperforms the previous solution in terms of delay.**

***Index Terms*—Admission control, resource allocation, utility maximization.**

I. Introduction

Demand for high data rate has been sharply increasing with the advent of smart devices and cloud computing. MaxWeight-based link rate allocation scheme of [1] has attracted great interest, since it achieves the optimal throughput in communication networks without a prior knowledge of traffic arrivals, and can play a core role in attaining the maximum utility in the Network Utility Maximization (NUM) framework [2]. However, the MaxWeight-based schemes are highly complex, i.e., NP-Hard in general [3], and under input traffic out of the capacity region, they fail to control the network system for the maximum utility [4].

It has been shown in [4] that an admission control can be used jointly to achieve the maximum utility for *any* input traffic. The solution, however, requires each source node to manage more state variables, adding additional complexity to the already complex MaxWeight-based algorithm. Further, these state variables are non-intuitive and their impact on the performance is not even fully understood. In this letter, we develop a simpler joint solution that provably achieves the maximum utility without additional state information, as well as empirically achieves better delay performance.

II. System model

We consider a network graph $G = (N, L)$, where N is the set of nodes and L is the set of directed links. We assume that the time is slotted, and simultaneous link transmissions can interfere each other. For ease of exposition, we assume fixed transmission power, fixed link rate if the transmission does not interfere with others, and single-hop traffic, where the packets transmitted over a link immediately leave the system. Extension to multiple power levels, time-varying link rates, and multi-hop traffic is straightforward. We refer interested readers to [4], [5].

Let $\tilde{A}_l(t)$ denote the external arrivals to link l at time t. We assume that it is bounded by $\tilde{A}_l(t) < \tilde{A}_l^{max}$ and has an expectation $\mathbf{E}\big[\tilde{A}_l(t)\big] = \lambda_l$ regardless of t. Among $\tilde{A}_l(t)$, $A_l(t)$ is accepted to link l by admission control scheme at time t. Each link l, if it transmits successfully without interfering other concurrent transmissions, has a fixed transmission rate of $R_l(t)$ at time t, which is also bounded by $R_l(t) < R_l^{max}$. Let $Q_l(t)$ denote the queue length of link l. At each time t, it evolves as

$$Q_l(t+1) = \max\big(Q_l(t) - R_l(t), 0\big) + A_l(t). \qquad (1)$$

We assume the infinite queue size. Thus the traffic is not dropped in the queue once it is accepted by the admission controller.

A *feasible* schedule is a set of links such that all the links can transmit simultaneously without interfering each other. A feasible schedule can be represented by a (feasible) link rate vector, where each element for active link l is $R_l(t)$ and each for inactive link l is 0. Let Γ denote the set of all the feasible link rate vectors. It has been shown that the convex hull of Γ is the capacity region Λ of the network [1].

We denote the utility of the session at link l when it is served at rate a_l by $U_l(a_l)$. We here consider the elastic traffic only. Thus the utility function is assumed to be a monotonically increasing concave function of a_l, satisfying $U_l(a_l) \geq 0$. Let $\mathbf{a}^* = (a_l^*)$ denote an optimal rate vector that maximizes the total utility, i.e.,

$$\mathbf{a}^* := \operatorname*{argmax}_{(a_l)\in\Lambda, a_l \leq \lambda_l} \sum_l U_l(a_l).$$

Our objective is to design a joint admission control and link rate allocation (or scheduling) scheme that attains the maximum total utility while managing the network stable for an arbitrary incoming traffic. A network is said to be *stable* if the time average of the expected total queue length remains finite, i.e., $\lim_{\tau\to\infty} \frac{1}{\tau}\sum_{t=0}^{\tau-1} \mathbf{E}\big[\sum_l Q_l(t)\big] < \infty$.

III. Previous Joint Solution

In [4], it has been shown that a joint solution of admission control and link scheduling can solve the problem. While the

Manuscript received July 19, 2012. The associate editor coordinating the review of this letter and approving it for publication was H. Luo.

J. Choi is with the Department of Information and Communication Engineering, Yeungnam University, Korea (e-mail: jchoi@yu.ac.kr), W. Pak is with Jangwee Research Institute for National Defence, Ajou University, Suwon, Korea (e-mail: wgpak@ajou.ac.kr), and C. Joo, the corresponding author, is with the School of Electrical and Comupter Engineering, Ulsan National Institute of Science and Technology, Korea (e-mail: cjoo@unist.ac.kr).

Digital Object Identifier xx.xxxx/LCOMM.2012.xxxxxx.xxxxxx

previous solution has been developed for multi-hop sessions, we provide its single-hop-session counterpart for ease of exposition.

In their solution, each link l has a reservoir of maximum size B_l^{max} ahead of its queue. The reservoir keeps $B_l(t)$ bits at time t. Among the bits backlogged in the reservoir and new arrivals at time t, the admission controller accepts $A_l(t)$ bits into the queue. The rest bits are saved at the reservoir for the next slots. Hence, the backlog $B_l(t)$ evolves as $B_l(t+1) = \min\big(B_l(t) + \tilde{A}_l(t) - A_l(t), B_l^{max}\big)$, and the accepted data $A_l(t)$ is bounded by $A_l^{max} := B_l^{max} + \tilde{A}_l^{max}$.

The solution operates as follows [4].

- **Admission control:** The accepted input traffic, $A_l(t)$ is determined as

$$A_l(t) = \underset{0 \le a_l \le B_l(t) + \tilde{A}_l(t)}{\operatorname{argmax}} \; a_l(\eta Z_l(t) - Q_l(t)),$$

  where $Z_l(t)$ is a queue-like state variable traced by the admission controller, and $0 < \eta \le 1$ is a constant. The detailed behavior of state variable is as follows.

$$Z_l(t+1) = \max\big(Z_l(t) - S_l(t), 0\big) + A_l^{max} - A_l(t),$$

  where its service rate $S_l(t)$ is selected as

$$S_l(t) = \underset{0 \le s_l \le A_l^{max}}{\operatorname{argmax}} \; \big(s_l \cdot \eta Z_l(t) + K U_l(A_l^{max} - s_l)\big)$$

  for some constant $K(>0)$.
- **Link scheduling:** The link scheduler chooses a set of simultaneous transmissions, i.e., a feasible link rate vector, by solving the Maximum Weighted Matching (MWM) problem.

$$\mathbf{R}(t) = \underset{(r_l) \in \Gamma}{\operatorname{argmax}} \sum_l r_l Q_l(t). \tag{2}$$

Under this solution, the total utility approaches the maximum as $K \to \infty$ even when the external arrival rate vector is out of the capacity region [4].

IV. Proposed Joint Solution

The previous solution has auxiliary state variables $Z_l(t)$ and their corresponding services $S_l(t)$, whose meaning and impact on the system performance, however, are not fully understood. In this section, we show that these state parameters can be removed without sacrificing the total utility even for the traffic out of the capacity region. Our admission control scheme directly controls the external arrivals and does not require any additional structure like the reservoir. If a reservoir is required, e.g., for traffic shaping, it can be place in front of the link queue. Our proposed scheme will consider the output from the reservoir as the external arrivals.

We begin with the description of our solution.

- **Admission control:** The accepted input traffic is

$$A_l(t) = \underset{0 \le a_l \le \tilde{A}_l(t)}{\operatorname{argmax}} \; \left(K U_l(a_l) - a_l Q_l(t)\right), \tag{3}$$

  where $K(>0)$ is some constant.
- **Link scheduling:** It solves the MWM problem of (2).

We show that with our scheme, the network is stable for any external arrivals. First we define a set of rate vectors Λ_ϵ, which is a reduced set from the capacity region by $\epsilon(>0)$, i.e., $\Lambda_\epsilon = \{(a_l) \mid (a_l + \epsilon) \in \Lambda,\ a_l \ge 0,\ \forall l\}$. Let $\mathbf{a}^*(\epsilon)$ denote the link rate vector in the ϵ-reduced capacity region with the maximum utility, i.e., $\mathbf{a}^*(\epsilon) = \operatorname{argmax}_{(a_l) \in \Lambda_\epsilon, (a_l) \le (\lambda_l)} \sum_l U_l(a_l)$. Clearly, we have $\Lambda_\epsilon \to \Lambda$ and $\mathbf{a}^*(\epsilon) \to \mathbf{a}^*$ as $\epsilon \to 0$.

We borrow the following two lemmas from [4], which are important in the performance analysis of our proposed scheme. The first lemma specifies a property of the link scheduling algorithm.

Lemma 1: Link scheduler (2) satisfies

$$\sum_l \mathbf{E}\big[R_l(t) Q_l(t) | \mathbf{Q}(t)\big] \ge \sum_l (a_l^*(\epsilon) + \epsilon) Q_l(t)$$

for small $\epsilon(>0)$

The second lemma is a variant of Foster's stability criterion.

Lemma 2: A countable Markov chain $\{\mathbf{Q}(t)\}_{t=0}^{\infty}$ is positive recurrent if there exists a Lyapunov function $V(\mathbf{Q}(t)) \ge 0$ and a finite set of states F such that

i) if $\mathbf{Q}(t) \in F$, $\Delta(\mathbf{Q}(t)) < \infty$, and
ii) if $\mathbf{Q}(t) \notin F$, $\Delta(\mathbf{Q}(t)) < -\delta$ for a positive constant δ,

where the conditional Lyapunov drift is defined as $\Delta(\mathbf{Q}(t)) := \mathbf{E}\big[V(\mathbf{Q}(t+1)) - V(\mathbf{Q}(t)) | \mathbf{Q}(t)\big]$.

It is well-known that the queue lengths have a finite expectation and thus, the network is stable if the underlying Markov chain is a positive recurrent. Hence we prove the network stability under our solution by showing that there exists a Lyapunov function with negative drift for sufficiently large queue lengths.

We define the quadratic Lyapunov function

$$V(\mathbf{Q}(t)) := \frac{1}{2} \sum_l (Q_l(t))^2,$$

whose conditional drift $\Delta(\mathbf{Q}(t))$ is bounded, from (1), as

$$\begin{aligned} \Delta(\mathbf{Q}(t)) \le & -\sum_l \mathbf{E}[R_l(t) Q_l(t) | \mathbf{Q}(t)] \\ & + \sum_l \mathbf{E}\left[A_l(t) Q_l(t) | \mathbf{Q}(t)\right] + C_1 \end{aligned} \tag{4}$$

where $C_1 := \frac{1}{2} \sum_l \big((R_l^{max})^2 + (\tilde{A}_l^{max})^2\big)$.

We here consider a *nominal* admission control. This nominal admission controller is not very practical since it requires the knowledge of the arrival rate λ_l and the optimal link rate vector $\mathbf{a}^*(\epsilon)$ in the ϵ-reduced capacity region. We just use this as a reference scheme to show the optimal performance of our admission controller. Under the nominal admission controller, each link l accepts $a_l^*(\epsilon)\lambda_l^{-1}$ fraction of the arrivals regardless of the queue length, i.e., the accepted traffic is $A_l^{nom}(t) = a_l^*(\epsilon)\lambda_l^{-1}\tilde{A}_l(t)$. The expectation is $\mathbf{E}\big[A_l^{nom}(t)\big] = a_l^*(\epsilon)$ then.

We evaluate the total utility of our admission control in comparison with the nominal admission control. From (3), we have

$$K U_l(A_l(t)) - A_l(t) Q_l(t) \ge K U(A_l^{nom}(t)) - A_l^{nom}(t) Q_l(t). \tag{5}$$

Rearranging the inequality, we obtain

$$A_l(t) Q_l(t) \le A_l^{nom}(t) Q_l(t) + K(U_l(A_l(t)) - U_l(A_l^{nom}(t))).$$

Taking conditional expectation on both sides and summing over all the links, we have

$$\begin{aligned}\sum_l \mathbf{E}\left[A_l(t)Q_l(t)|\mathbf{Q}(t)\right] &\le \sum_l \mathbf{E}\left[A_l^{nom}(t)Q_l(t)|\mathbf{Q}(t)\right] + C_2 \\ &= \sum_l a_l^*(\epsilon)Q_l(t) + C_2,\end{aligned} \tag{6}$$

where we defined a constant $C_2 := K\sum_l U_l(A_l^{max})$ which is no smaller than $K\sum_l\left(U_l(A_l(t)) - U_l(A_l^{nom}(t))\right)$. The equality comes from $\mathbf{E}[A_l^{nom}(t)Q_l(t)|\mathbf{Q}(t)] = \mathbf{E}[A_l^{nom}(t)]Q_l(t) = a_l^*(\epsilon)Q_l(t)$.

By applying Lemma 1 and (6) to (4), we have

$$\begin{aligned}\Delta(\mathbf{Q}(t)) &\le -\sum_l (a_l^*(\epsilon)+\epsilon)Q_l(t) \\ &+\Big(\sum_l a_l^*(\epsilon)Q_l(t) + C_2\Big) + C_1 \\ &= -\epsilon\sum_l Q_l(t) + C,\end{aligned} \tag{7}$$

where $C := C_1 + C_2$. Thus, the Lyapunov function has a negative drift whenever the queue length is sufficiently large. According to Lemma 2 with $F := \{\mathbf{Q} \mid \sum_l Q_l > \epsilon^{-1}(C+\delta)\}$ and any $\delta > 0$, we can conclude that the underlying Markov chain is positive recurrent and hence, the network is stable. Actually, the average total queue length is bounded as follows [5]. Let us take the expectation on both sides of (7) over the distribution of $\mathbf{Q}(t)$, which leads to $\mathbf{E}[V(\mathbf{Q}(t+1)) - V(\mathbf{Q}(t))] \le -\epsilon\mathbf{E}\Big[\sum_l Q_l(t)\Big] + C$. We sum up the inequality from $t = 0$ to $\tau - 1$ and divide by τ. Then, $\frac{1}{\tau}(\mathbf{E}[V(\mathbf{Q}(\tau))] - \mathbf{E}[V(\mathbf{Q}(0))]) \le -\epsilon\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbf{E}\Big[\sum_l Q_l(t)\Big] + C$. Finally, letting $\tau\to\infty$, we obtain $\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbf{E}\Big[\sum_l Q_l(t)\Big] \le \frac{C}{\epsilon}$ since $\mathbf{E}[V(\mathbf{Q}(0))]$ is finite.

Henceforth we show that under our proposed scheme, the total utility approaches the maximum as closely as we desire. We derive the result starting from (4). By adding and subtracting $\sum_l \mathbf{E}[KU_l(A_l(t))|\mathbf{Q}(t)]$ to the right side of (4) and using (5), we have

$$\begin{aligned}\Delta(\mathbf{Q}(t)) &\le -\sum_l \mathbf{E}[R_l(t)Q_l(t)|\mathbf{Q}(t)] \\ &+\sum_l \mathbf{E}[A_l^{nom}(t)Q_l(t) - KU_l(A_l^{nom}(t))|\mathbf{Q}(t)] \\ &+K\sum_l \mathbf{E}[U_l(A_l(t))|\mathbf{Q}(t)] + C_1.\end{aligned} \tag{8}$$

We notice the following inequality here:

$$\mathbf{E}[U_l(A_l^{nom}(t))] \ge U_l(\mathbf{E}[A_l^{nom}(t)]) - U_l(\tilde{A}_l^{max})$$

since $U_l(\mathbf{E}\big[A_l^{nom}(t)\big]) \le U_l(\tilde{A}_l^{max}) \le U_l(\tilde{A}_l^{max}) + \mathbf{E}\big[U_l(A_l^{nom}(t))\big]$. Combining this with (8), and applying Lemma 1 and $\mathbf{E}[A_l^{nom}(t)Q_l(t)|\mathbf{Q}(t)] = a_l^*(\epsilon)Q_l(t)$, we obtain

$$\begin{aligned}\Delta(\mathbf{Q}(t)) &\le -\epsilon\sum_l Q_l(t) \\ &\quad - K\Big(\sum_l U_l(\mathbf{E}[A_l^{nom}(t)]) - \sum_l U_l(\tilde{A}_l^{max})\Big) \\ &\quad + K\sum_l \mathbf{E}[U_l(A_l(t))|\mathbf{Q}(t)] + C_1 \\ &= -\epsilon\sum_l Q_l(t) - K\sum_l U_l(a_l^*(\epsilon)) \\ &\quad + K\sum_l \mathbf{E}\big[U_l(A_l(t))|\mathbf{Q}(t)\big] + D,\end{aligned} \tag{9}$$

where $D := C_1 + K\sum_l U_l(\tilde{A}_l^{max})$.

Next, we take the expectation on the both sides of (9) over the distribution of $\mathbf{Q}(t)$ and then

$$\begin{aligned}\mathbf{E}\big[V(\mathbf{Q}(t+1)) - V(\mathbf{Q}(t))\big] &\le -\epsilon\sum_l \mathbf{E}\big[Q_l(t)\big] \\ &- K\sum_l U_l(a_l^*(\epsilon)) + K\sum_l \mathbf{E}\big[U_l(A_l(t))\big] + D.\end{aligned}$$

We sum up the above equations from $t = 0$ to $\tau-1$ and divide by $K\tau$. By letting $\tau\to\infty$, we can obtain

$$\begin{aligned}\sum_l \mathbf{E}\big[U_l(A_l(t))\big] &\ge \sum_l U_l(a_l^*(\epsilon)) \\ &+ \frac{1}{K}\Big(\epsilon\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbf{E}\Big[\sum_l Q_l(t)\Big] - D\Big).\end{aligned} \tag{10}$$

Both $\mathbf{E}\big[V(\mathbf{Q}(\tau))\big]$ and $\mathbf{E}\big[V(\mathbf{Q}(0))\big]$ disappear since they are finite. Also recall that $\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbf{E}\big[\sum_l Q_l(t)\big] < \infty$ since the network is stable under our scheme. From the concavity of utility function, we can conclude that the total utility approaches the maximum as $K\to\infty$ and $\epsilon\to 0$, i.e.,

$$\sum_l U_l(\mathbf{E}\big[A_l(t)\big]) \ge \sum_l \mathbf{E}\big[U_l(A_l(t))\big] \to \sum_l U_l(a_l^*).$$

V. Simulations

We evaluate our joint solution through simulations. We consider a wireless network of Fig. 1, where 12 nodes are randomly placed in the 1000×1000 square space and two nodes are connected by an unit-bandwidth link if their distance is no greater than 400. We assume the primary interference model, under which two links that share a node cannot be activated simultaneously. We establish a session on each link and select its direction at random, which is shown as an arrow in Fig. 1. At every time slot, data bits arrive at each link following Geometric distribution truncated at 1000 (units).

We evaluate our solution in comparison with the previous state-of-the-art shown in Section III and a solution without any admission control. We set $K = 10.0$ and $U_l(a) = a$ for all the the links l for the schemes, and set the additional parameters of the previous solution to $B_l^{max} = 10.0$ and $\eta = 0.1$. We obtain our results by running the simulations for 10^6 time slots.

By increasing the mean arrival rate from 0 to 2.0 (units/slot) for all the sessions, we measure average queue lengths of sessions and total utility. The results are shown in Fig. 2. It

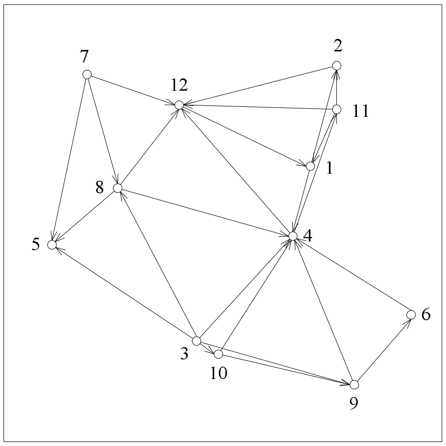

Fig. 1. Network topology for simulations.

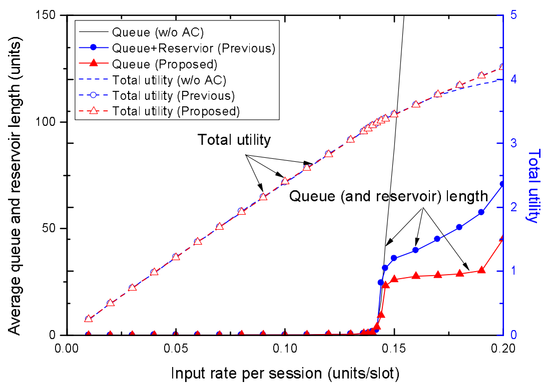

Fig. 2. Average queue length and total utility.

shows that without admission control, the queue size builds up at the input rate of around 1.5, which is the boundary of the capacity region. In contrast, under both the previous and the proposed solutions, the queue lengths remain finite even for the input rates beyond the boundary. For the previous solution, we have added the reservoir backlogs to the queue length since it is directly related to delay performance. Our results show that the proposed solution notably outperforms the previous solution in terms of the queue length, i.e., delay. For utility, the achieved total utilities are close except the solution without admission controls, which achieves smaller utilities especially in heavy traffic region. As expected, the previous and the proposed solution show similar performance in terms of utility since they achieve the maximum utility.

Acknowledgment

This research was supported by the Yeungnam University research grants in 2011 and the Basic Science Research Program through the NRF funded by the MEST (No. 2012-0003227).

References

[1] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936-1948, Dec. 1992.

[2] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452-1463, Aug. 2006.

[3] C. Joo, G. Sharma, N. Shroff, and R. Mazumdar, "On the Complexity of Scheduling in Wireless Networks," *EURASIP Journal of Wireless Communications and Networking*, Oct. 2010.

[4] M. Neely, E. Modiano, and C. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 2, pp. 396-409, Apr. 2008.

[5] M. Neely, E. Modiano, and C. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 89-103, Jan. 2005.