Network Utility Maximization with Heterogeneous Traffic Flows

Abhishek Sinha, Eytan Modiano
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139
Email: sinhaa@mit.edu, modiano@mit.edu

***Abstract*— We consider the Network Utility Maximization (NUM) problem for wireless networks in the presence of arbitrary types of flows, including unicast, broadcast, multicast, and anycast traffic. Building upon the recent framework of a universal control policy (UMW), we design a utility optimal cross-layer admission control, routing and scheduling policy, called UMW+. The UMW+ policy takes packet level actions based on a *precedence-relaxed* virtual network. Using Lyapunov optimization techniques, we show that UMW+ maximizes network utility, while simultaneously keeping the physical queues in the network stable. Extensive simulation results validate the performance of UMW+; demonstrating both optimal utility performance and bounded average queue occupancy. Moreover, we establish a precise one-to-one correspondence between the dynamics of the virtual queues under the UMW+ policy, and the dynamics of the dual variables of an associated offline NUM program, under a subgradient algorithm. This correspondence sheds further insight into our understanding of UMW+.**

I. Introduction

Increasingly, networks need to support a heterogeneous mix of traffic that includes unicast, multicast and broadcast traffic flows. Such traffic may be comprised of video streaming, file downloads, distributed computation and storage, as well as a host of other applications. The increase in both volume, as well as QoS requirements, of these emerging applications can significantly strain network resources, especially in bandwidth limited wireless networks. This calls for efficient resource allocation schemes for wireless networks with heterogeneous type of traffic demands. Unfortunately, however, existing resource allocation schemes, both in theory and practice, are limited to dealing with point-to-point unicast traffic. While there has been some work on resource allocation for multicast and broadcast flows, until very recently there has been no mechanism for allocating resources in networks with arbitrary traffic flows that include concurrent unicast, multicast and broadcast traffic. In [1], the authors developed the first such algorithm for joint routing and scheduling in wireless networks with general traffic flows. The algorithm of [1], called *Universal Max Weight is guaranteed to stabilize the network for any arrivals that are* within the network capacity (stability) region. However, it provides no service guarantees for traffic that is outside the network capacity region.

When traffic arrivals are outside of the network's capacity region, an admission control mechanism is needed to limit the admitted traffic to being within the network's capacity region. In particular, a resource allocation scheme is needed that can optimize user-level performance across the network. Network Utility Maximization (NUM) is an approach for allocating resources across the network in a manner that maximizes overall "utility," where user's level of satisfaction is measured using a concave function of the allocated data rate to that user (i.e., the utility function).

Starting with the work of Kelly nearly two decades ago, [2], there has been a tremendous amount of work on the NUM problem (see, for example, [3], [4], [5] and references therein). These approaches use convex optimization and Lagrangian duality to optimize overall network utility in networks with static traffic demand. In the context of stochastic traffic, the works of [6], [7], [8] used Lyapunov optimization techniques to develop utility optimal flow control and resource allocation for wireless networks with unicast traffic demands. Unfortunately, however, all of these works deal with unicast (point-to-point) traffic demands.

In this paper, we develop a jointly optimal admission control, routing, and scheduling algorithm for networks with generalized traffic flows. Our approach builds upon the Universal Max Weight (UMW) algorithm of [1] that solves the optimal routing and scheduling problem for networks with general traffic flows, thus we call our algorithm *UMW+*. To the best of our knowledge, UMW+ is the first efficient solution to the NUM problem for networks with generalized traffic flows, including, but not limited to concurrent unicast, broadcast, multicast, and anycast traffic. We derive UMW+ by using the concept of *precedence-relaxed* virtual queueing network [1]. Moreover, by formulating the NUM problem as a static convex optimization problem, we show that, under the dual subgradient algorithm, the evolution of the dual variables precisely corresponds to the evolution of the virtual queues in UMW+. This sheds a new light on the relationship between Max-Weight and sub-gradient algorithms [9]. We make the following key contributions in this paper:

- We design the first efficient policy for the NUM problem for *arbitrary* types of traffic, including concurrent unicast, broadcast, multicast, and anycast sessions.
- The policy was derived by making effective use of the *virtual network* framework, obtained by relaxing the precedence constraints in a network [1]. This methodology was also employed earlier in a wireless broadcasting problem [10] and a distributed function computation problem [11].

This work was supported by the NSF Grant CNS-1524317, and DARPA I2O and Raytheon BBN Technologies under Contract No. HROO II-I 5-C-0097.

- We formulate a static version of the NUM problem by decomposing the flows *route-wise*. This is to be contrasted with earlier works on this problem (*e.g.*, [4], [5]), which use node-based flow conservation, applicable *only* to the unicast flows. We show that, under the dual subgradient algorithm with unit step size, the evolution of the dual variables *exactly* corresponds to the evolution of the virtual queues under the proposed UMW+ policy.
- We explicitly characterize the dual objective function of the above optimization problem in terms of cost of the shortest route for each source and the Fenchel conjugates of the utility function.
- Finally, we validate our theoretical results through extensive numerical simulations.

The rest of the paper is organized as follows: In Section II, we describe the network and traffic model. In Section III, we detail the virtual network framework and derive the UMW+ control policy. In Section IV, we prove the stability of the physical queues under the UMW+ policy. In Section V, we establish the equivalence between the UMW+ policy and the dual subgradient algorithm. In Section VI, we provide simulation results and conclude the paper in Section VII.

II. System Model and Problem Formulation

A. System Model

Network Model: Let the topology of a wireless network be given by the graph $\mathcal{G}(V,E)$ - where V is the set of n nodes and E is the set of m links. Time is slotted, and at every slot, each wireless link could be in either the ON (1) or, OFF (0) state. The random link state process $\boldsymbol{\sigma}(t) \in \{0,1\}^m$ is assumed to be evolving according to a stationary ergodic stochastic process. If at a slot t, link e is ON and is *activated*, it can transmit c_e packets in that slot [1].

Interference Model: Due to the wireless inter-channel interference, not all links can be activated at a slot simultaneously. The set of all interference-free links, which may be activated together in a slot, is given by $\mathcal{M} \subseteq 2^E$. As an example, in the case of *primary* interference constraints, the set $\mathcal{M}$ consists of the set of all *Matchings* in the graph. On the other hand, in an interference-free wired network, the set $\mathcal{M}$ consists of the set of all subsets of links.

Traffic and Utility Model: External packets from different traffic *classes* are admitted to the network by an admission controller $\mathcal{A}$. The set of all possible classes of traffic is denoted by $\mathcal{K}$. A traffic class $k \in \mathcal{K}$ is associated with the following two attributes - (1) type of the traffic (*e.g.*, unicast, broadcast, multicast or anycast), and (2) a monotone increasing strictly concave Utility function $U_{(k)} : \mathbb{R}_+ \to \mathbb{R}_+$. Let $\mathcal{T}^k$ denote the set of all possible routes for routing packets from class k in the graph $\mathcal{G}$. As an example, for a unicast class k with source node s and destination node t, the set $\mathcal{T}^k$ consists of the set of all $s-t$ paths in the graph. Similarly, for a broadcast class, the set of all routes is given by the set of all spanning trees in the graph. We consider an *infinitely backlogged* traffic model, where the admission controller $\mathcal{A}$ has potentially an unlimited number of packets from each class available for admission.

B. Admissible Policies

An admissible policy π consists of the following three modules: (1) an admission controller $\mathcal{A}$, (2) a routing module $\mathcal{R}$, and (3) a link scheduler module $\mathcal{S}$. The admission controller determines the number of external packets to be admitted to the network from each class in each slot. We assume that, due to physical constraints (power, capacity limitations etc.), at most $A_{\max}$ number of packets may be feasibly admitted from any class per slot [2]. The routing module routes the admitted packets according to the type of flow they belong to, and the link scheduler activates a subset of interference-free links from the set $\mathcal{M}$ in every slot. The set of all admissible policies is denoted by Π.

C. The Network Utility Maximization (NUM) Problem

The Network Utility Maximization problem seeks to find an admissible policy $\pi \in \Pi$, which maximizes the sum utility of all classes, while keeping the queues in the network *stable* [3]. Formally, let the random variable $R_k^\pi(T)$ denote the number of packets received in common by the destination(s) of class k up to time T under the action of an admissible policy π. Denote the random queue length of packets waiting to cross the edge e under the policy π by $Q_e^\pi(T)$. The NUM problem seeks to find a policy $\pi^* \in \Pi$, which solves the following:

$$\max_{\pi \in \Pi} \mathbb{E} \sum_k U_k(r_k) \tag{1}$$

Subject to,

$$\lim_{T\to\infty} \frac{R_k^\pi(T)}{T} = r_k, \ \ \forall k \in \mathcal{K}, \ \text{w.p. 1.,} \tag{2}$$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{e \in E} Q_e^\pi(T) = 0, \quad \text{w.p. 1.,} \tag{3}$$

where the expectation and the almost sure limits are taken over the randomness of the network configurations $\boldsymbol{\sigma}(t)$ and possible randomness in the policy. In the following, we will drop the superscript π from the random variables when the driving policy π is clear from the context.

III. The Virtual Network Framework

In this section, we describe a *virtual network* framework, obtained by relaxing the natural precedence constraints associated with a multi-hop network. Our goal in this section is to design a utility-optimal stabilizing control policy for the simpler *virtual queueing system*. Section IV shows that when the same policy is used in the actual physical network, the physical queues are also stable. The virtual network methodology in a similar context was first introduced in [1]. Consider an admissible policy π, which admits $A^k(t)$ packets

[1]A link does not transmit any packet if it is OFF.

[2]Taking $A_{\max} \equiv \sum_e c_e$ yields the same optimal utility obtained without this constraint.

[3]Throughout the paper, by stability, we will mean almost sure rate stability, defined in Eqn. (3).

from class k, and activates the links $\boldsymbol{\mu}(t) \in \{0,1\}^m$ at slot t. Taking the random wireless link states into account, the service rate for the virtual queue $\tilde{Q}_e$ is $\mu_e(t)c_e\sigma_e(t)$ packets per slot. Since the virtual network is precedence-relaxed, all admitted packets *immediately* enter all virtual queues on the selected route, i.e., unlike the physical network, arrivals to a virtual queue need not wait to cross the intermediate links [1]. Thus, the virtual queues $\tilde{\boldsymbol{Q}}(t)$ evolve as

$$\tilde{Q}_e(t+1) = \big(\tilde{Q}_e(t) + \tilde{A}_e(t) - \mu_e(t)c_e\sigma_e(t)\big)^+ \quad \forall e \in E, \quad (4)$$

where $\tilde{A}_e(t)$ is the total number of (controlled) arrival of packets to the virtual queue $\tilde{Q}_e$ at slot t. Clearly, $\tilde{A}_e(t)$ depends on the routes selected by the routing module $\mathcal{R}$, i.e.,

$$\tilde{A}_e(t) = \sum_k A^k(t)\mathbb{1}(e \in T^k(t)). \quad (5)$$

In order to design a utility-optimal stabilizing control policy for the virtual queues, we use the *drift-plus-penalty* framework of [12]. Consider the following Lyapunov function, which is quadratic in the virtual queue lengths (as opposed to the usual physical queue lengths [12])

$$L(\tilde{\boldsymbol{Q}}(t)) = \sum_e \tilde{Q}_e^2(t). \quad (6)$$

The one-slot conditional drift of $L(\tilde{\boldsymbol{Q}}(t))$ under the action of a control policy π is given as follows

$$\begin{aligned}
&\Delta(\tilde{\boldsymbol{Q}}(t),\boldsymbol{\sigma}(t)) \equiv \mathbb{E}\Big(L(\tilde{\boldsymbol{Q}}(t+1)) - L(\tilde{\boldsymbol{Q}}(t))|\tilde{\boldsymbol{Q}}(t),\boldsymbol{\sigma}(t)\Big)\\
&\overset{(a)}{\leq} \mathbb{E}\Big(\sum_e \big(\tilde{A}_e^2(t) + c_e^2\\
&+2\tilde{Q}_e(t)(\tilde{A}_e(t) - \mu_e(t)c_e\sigma_e(t))\big)|\tilde{\boldsymbol{Q}}(t),\boldsymbol{\sigma}(t)\Big)\\
&= B + 2\mathbb{E}\Big(\sum_e \tilde{Q}_e(t)\sum_k \big(A^k(t)\mathbb{1}(e \in T^k(t))\big)|\tilde{\boldsymbol{Q}}(t),\\
&\boldsymbol{\sigma}(t)\Big) - 2\mathbb{E}\Big(\sum_e \tilde{Q}_e(t)\mu_e(t)c_e\sigma_e(t)|\tilde{\boldsymbol{Q}}(t),\boldsymbol{\sigma}(t)\Big)\\
&\overset{(b)}{=} B + 2\mathbb{E}\Big(\sum_k A^k(t)\big(\sum_e \tilde{Q}_e(t)\mathbb{1}(e \in T^k(t))\big)|\tilde{\boldsymbol{Q}}(t),\\
&\boldsymbol{\sigma}(t)\Big) - 2\mathbb{E}\Big(\sum_e \tilde{Q}_e(t)\mu_e(t)c_e\sigma_e(t)|\tilde{\boldsymbol{Q}}(t),\boldsymbol{\sigma}(t)\Big), \quad (7)
\end{aligned}$$

where the inequality (a) is obtained by using the virtual queue dynamics in Eqn. (4), the equality (b) is obtained by interchanging the order of summation in the first term, B is a finite constant, upper bounded by $kmA_{\max}^2 + mc_{\max}^2$, where $A_{\max}$ is the maximum number of external admissions per slot per class (defined in Section II-B) and $c_{\max} \overset{\text{def}}{=} \max_{e\in E} c_e$. Moreover, admission of $A^k(t)$ packets from class k in slot t yields a "one-slot utility" of $U_k(A^k(t))$ for class k. Following the *"Drift-Plus-Penalty"* framework of [12], we consider a cross-layer admission control, routing and link scheduling policy, which is obtained by minimizing the objective function $(*)$ given at the bottom of this page, over all admissible controls $(\boldsymbol{A}(t), \boldsymbol{T}(t), \boldsymbol{\mu}(t))$ per slot. For the objective $(*)$, we have substituted the upper bound of the drift $\Delta(\tilde{\boldsymbol{Q}}(t),\boldsymbol{\sigma}(t))$ from Eqn. (7) (without the constant B), and V is taken to be a fixed positive constant. This yields the following joint routing, admission control, and link scheduling policy π^*, which we call *Universal Max-Weight Plus* (**UMW+**):

1. Routing ($\mathcal{R}$)**:** The routing policy follows by minimizing the term (a) of the objective $(*)$. Consider a weighted graph $\tilde{\mathcal{G}}$, where each edge e is weighted by the corresponding virtual queue length $\tilde{Q}_e(t)$. Under the policy π^*, all admitted packets from class $k \in \mathcal{K}$ (refer to $\mathcal{A}$ in Part (2) below) are assigned a route corresponding to the shortest route $T^k(t) \in \mathcal{T}^k$ in $\tilde{\mathcal{G}}$. In particular,

- For a unicast $s{-}t$ flow, $T^k(t)$ is the weighted shortest $s-t$ path in $\tilde{\mathcal{G}}$.
- For a broadcast flow originating from the node r, $T^k(t)$ is the Minimum Weight Spanning Tree (MST) in the weighted graph $\tilde{\mathcal{G}}$.
- For a multicast flow, $T^k(t)$ is the corresponding Steiner tree in the weighted graph $\tilde{\mathcal{G}}$.
- For an anycast flow [13], $T^k(t)$ is the weighted shortest path from the source to *any* of the given destinations.

Denote the cost of the weighted shortest route corresponding to class k obtained above by $C^k(t)$, i.e.,

$$C^k(t) = \min_{T^k \in \mathcal{T}^k} \sum_e \tilde{Q}_e(t)\mathbb{1}(e \in T^k) \quad (8)$$

2. Admission Control ($\mathcal{A}$)**:** The admission control policy follows by jointly considering the terms (a) and (c) of the per-slot objective function $(*)$. The **UMW+** policy admits $A^k(t)$ packets from class $k \in \mathcal{K}$ in slot t, where $A^k(t)$ is obtained by the solution to the following one dimensional convex optimization problem:

$$A^k(t) = \underset{0\leq x\leq A_{\max}}{\arg\min} \big(C_k(t)x - VU_k(x)\big) \quad (9)$$

$$\textbf{OBJ.} = \underbrace{2\mathbb{E}\Big(\sum_k A^k(t)\big(\sum_e \tilde{Q}_e(t)\mathbb{1}(e \in T^k(t))\big)|\tilde{\boldsymbol{Q}}(t),\boldsymbol{\sigma}(t)\Big)}_{(a)} \underbrace{- 2\mathbb{E}\Big(\sum_e \tilde{Q}_e(t)\mu_e(t)c_e\sigma_e(t)|\tilde{\boldsymbol{Q}}(t),\boldsymbol{\sigma}(t)\Big)}_{(b)} \underbrace{- 2V\sum_k U_k(A^k(t))}_{(c)}, \quad (*)$$

In the case of differentiable utility functions $U_k(\cdot)$, the number of packets $A^k(t)$ admitted may be obtained in closed form as follows:

$$A^k(t) = [U_k'^{-1}(C_k(t)/V)]_0^{A_{\max}}, \qquad (10)$$

where $[\cdot]_a^b$ denotes projection onto the interval $[a, b]$.

3. Link Scheduling ($\mathcal{S}$): The drift-minimizing link scheduling policy is obtained by minimizing the term (b) of the per-slot objective function $(*)$. Consider the weighted graph $\hat{\mathcal{G}}$, where the link e of the graph $\mathcal{G}$ is given a weight of $c_e\tilde{Q}_e(t)$, if it is ON in slot t (i.e. $\sigma_e(t) = 1$), and 0, otherwise. Then the link scheduler $\mathcal{S}$ activates the set of ON links which maximizes the total weight among the set of all interference-free links $\mathcal{M}$, i.e.,

$$\boldsymbol{\mu}^*(t) \in \arg\max_{\boldsymbol{\mu}\in\mathcal{M}} \sum_e \tilde{Q}_e(t)c_e\mu_e\sigma_e(t). \qquad (11)$$

In the special case of primary interference constraints, where the set $\mathcal{M}$ consists of the set of all *Matchings* of the graph, the problem (11) corresponds to the Maximum-Weighted Matching problem, which can be solved efficiently even in a distributed fashion [14].
Our main result in this section is Theorem 1, which claims that under the UMW+ control policy described above, the virtual queues are stable and the average expected utility obtained is arbitrarily close to the optimal utility.

Theorem 1. *Let U^* be the optimal utility for the NUM problem given in Eqn.* (1)*. Then, under the action of the* ***UMW+*** *control policy, (a) the virtual queues are rate stable (in the sense of Eqn.*(3)*), and (b) the utility achieved is at least $U^* - \mathcal{O}(\frac{1}{V})$, for any $V > 0$.*

Proof. See the Appendix (Section VIII) of the tech-report [15]. □

An alternative treatment of Utility Optimality and Stability of the Virtual Queues under the UMW+ policy will be given in Section V, where we relate the virtual queue evolution to the subgradient descents of an appropriately defined dual optimization problem. This sheds new and fundamental insights in the construction and operation of the virtual queues.

IV. Control of the Physical Network

In the physical network, *the same* admission control, routing and link scheduling policy are used as in the virtual network of Section III. Hence, the same number of packets are admitted at each slot from each class to the virtual and the physical network. As a result, the utility achieved in these two networks are the same, however, their queueing evolutions are different. In this Section, we establish the stability of the physical queues, which, unlike the virtual queues, are subjected to the usual precedence constraint of a multi-hop network.

A. Packet arrivals to the Physical Queues

To connect the arrivals of packets in the virtual and the physical network, observe that since the routes are fixed at the sources, the total number of physical packets $A_e(t_1, t_2)$ admitted to the physical network in any time interval $(t_1, t_2]$, that *will cross edge* e *in future*, is the same as the number of virtual packets arrivals $\tilde{A}_e(t_1, t_2)$ to the corresponding virtual queue $\tilde{Q}_e$. Let the total service allocated to serve the virtual queue (resp. physical queue) in the time interval $(t_1, t_2]$ be denoted by $\tilde{S}_e(t_1, t_2)$ (resp. $S_e(t_1, t_2)$). Using the Skorokhod map [16] for the virtual queue iterations (4), we have

$$\tilde{Q}_e(t) = \sup_{0\leq t_1\leq t} \bigg(\tilde{A}_e(t_1, t) - \tilde{S}_e(t_1, t)\bigg). \qquad (12)$$

However, as noted above, we have

$$A_e(t_1, t_2) = \tilde{A}_e(t_1, t_2), S_e(t_1, t_2) = \tilde{S}_e(t_1, t_2). \qquad (13)$$

Hence, using Theorem (1) for the stability of the virtual queues and combining it with Eqns (12) and (13), we have

$$A_e(t_1, t) \leq S_e(t_1, t) + M(t), \forall t_1 < t, a.s., \qquad (14)$$

where $M(t) = o(t)$.

B. Stability of The Physical Queues

Note that, Theorem 1 establishes the stability of the *virtual queues* under the action of the **UMW+** policy. To prove the optimality of this policy in the NUM setting of Eqn. (1), we need to establish the stability of the physical queues $\{\boldsymbol{Q}(t)\}_{t\geq 1}$. This is a non-trivial task because, unlike the virtual queues, the dynamics of the physical queues is subject to the precedence constraints, and regulated by the *packet scheduling policy* employed in the physical network. A packet scheduling policy is a rule which resolves the contention when multiple packets want to cross the same edge at the same slot. The most common examples of packet scheduling policies include First In First Out (**FIFO**), Last In First Out (**LIFO**) disciplines etc. Following our earlier work [1], we consider a simple packet scheduling policy called the Extended Nearest to Origin (**ENTO**) policy:

Definition 1 (Extended Nearest to Origin policy)**.** *The ENTO policy prioritizes packets according to the decreasing order of their current distances (measured in hop-lengths) from their respective sources.*

For example, if there are two packets p_1 and p_2, which have traversed, say, 10 and 20 hops respectively from their sources, and wish to cross the same active edge e (with a unit capacity per slot) in the same slot, the ENTO policy prioritizes the packet p_1 over p_2 to cross the edge e.
Using ideas from adversarial queueing theory [17], it is shown in [1] (Theorem 3) that, under the ENTO packet scheduling policy, the almost sure packet arrival bound (14) also implies

the rate stability of the physical queues. Combining this result with Theorem 1, we conclude the following:

Theorem 2 (Stability of the Physical Queues)**.** *Under the action of the* ***UMW+*** *control policy, the physical queues are rate stable, i.e.,*

$$\lim_{T\to\infty} \frac{1}{T}\sum_{e\in E} Q_e(T) = 0, \quad \textit{w.p. 1.}$$

The proof of Theorem 2 follows directly from the proof of Theorem 3 of [1] and uses the bound (14) as a starting point. Combining Theorem 1 with Theorem 2, we conclude that the proposed **UMW+** policy is a utility optimal stabilizing control policy, which solves the NUM problem efficiently. This concludes the first half of the paper.

V. A Dual Perspective on the Virtual Queue Dynamics

In this Section, we consider the dual of an offline version of the NUM problem in a static wired network setting (no interference constraints). We give an alternative derivation of the UMW+ policy from an optimization theory perspective, which sheds further insight into the structure of the optimal policy. Our motivation in this section is similar to [5], [7], which give similar development for the Backpressure policy [18]. The dual problem is also of sufficient theoretical and practical interest, as the powerful machinery of convex optimization may be used to derive alternative efficient algorithms for solving the dual problem, which may then be translated to other dynamic policies (apart from the UMW+) for solving the NUM problem.

As in the previous section, the topological structure of a route depends on the type of flow - *e.g.*, a route is a path for unicast and anycast flows, a spanning tree for broadcast flows, a Steiner tree for multicast flows etc.

Fix a strictly positive constant V. Our goal is to solve the following utility maximization problem $\mathcal{P}$:

$$\text{Problem } \mathcal{P}: \quad \max V\sum_k U_k(r_k) \tag{15}$$

Subject to,

$$r_k = \sum_{p\in\mathcal{T}_k} f_p, \quad \forall k\in\mathcal{K}. \tag{16}$$

$$\sum_{p:e\in p} f_p \leq c_e, \quad \forall e\in E. \tag{17}$$

$$\boldsymbol{f} \geq \boldsymbol{0}. \tag{18}$$

The objective (15) denotes the total utility, scaled by V. The constraint (16) is obtained by decomposing the total incoming flow r_k to a class k into all available routes, where the variable f_p corresponds to the amount of flow carried by the route $p\in\mathcal{T}^k$. The constraint (17) corresponds to the capacity of edge e, and the constraint (18) corresponds to the non-negativity property of the flow variables.

In the special case of the NUM problem dealing exclusively with the unicast flows, the flow decomposition constraint (16) is usually replaced with the *flow conservation* constraints at the nodes, which leads to the Backpressure policy [5]. In contrast, we formulate the NUM problem with the flow decomposition constraint, since flow, in general, is not conserved at the nodes due to packet replications (as in broadcast and multicast flows).

A. The Dual Problem $\mathcal{P}^*$

The problem $\mathcal{P}$ is a concave maximization problem with linear constraints. To obtain its dual problem, we relax the capacity constraints (17) by associating a non-negative dual variable q_e with the constraint corresponding to the edge e. We choose *not to relax* the flow decomposition constraints (16) and the non-negativity constraints (18)[4]. This yields the following partial Lagrangian:

$$L(\boldsymbol{r},\boldsymbol{f},\boldsymbol{q}) = V\sum_k U_k(r_k) + \sum_e q_e(c_e - \sum_{p:e\in p} f_p), \tag{19}$$

leading to the following dual objective function $D(\boldsymbol{q})$:

$$D(\boldsymbol{q}) := \max_{\boldsymbol{r},\boldsymbol{f}\geq\boldsymbol{0}} L(\boldsymbol{r},\boldsymbol{f},\boldsymbol{q}), \tag{20}$$

Subject to,

$$r_k = \sum_{p\in\mathcal{T}_k} f_p, \quad \forall k\in\mathcal{K}. \tag{21}$$

By strong duality [19], the problem $\mathcal{P}$ is equivalent to the following dual problem:

$$\text{Problem } \mathcal{P}^*: \quad \min_{\boldsymbol{q}\geq\boldsymbol{0}} D(\boldsymbol{q}). \tag{22}$$

We next establish a simple lemma which will be useful for our subsequent development.

Lemma 1. *For any fixed $\boldsymbol{q}\geq\boldsymbol{0},\boldsymbol{r}\geq\boldsymbol{0}$, an optimal solution to the problem* (20) *is obtained by routing the entire flow r_k from each class k along a weighted shortest route $p_k^*\in\mathcal{T}^k$, weighted by the corresponding dual variables $\boldsymbol{q}$.*

Proof: Exchanging the order of summation in the last term of the Lagrangian in Eqn. (19), we have

$$L(\boldsymbol{r},\boldsymbol{f},\boldsymbol{q}) = V\sum_k U_k(r_k) + \sum_e q_e c_e - \sum_p f_p\big(\sum_{e:e\in p} q_e\big). \tag{23}$$

[4]Relaxation of these constraints yields a different dual problem.

Define $c_p(\boldsymbol{q}) \equiv \sum_{e:e\in p} q_e$ to be the cost of the route p where each edge e is weighted by the dual variable q_e, $\forall e \in E$. Hence, from the last term of the above expression (23) and the flow decomposition constraint (21), it immediately follows that the objective (20) is maximized by routing the entire incoming flow r_k along a weighted shortest route $p_k^* \in \mathcal{T}_k$ corresponding to class k. In other words, for any class k,

$$f_p^* = \begin{cases} r_k, & \text{if } p = p_k^* \\ 0, & \text{if } p \in \mathcal{T}^k \text{ and } p \neq p_k^*, \end{cases}$$

where $p_k^* = \arg\min_{p\in\mathcal{T}^k} c_p(\boldsymbol{q})$ (ties are broken arbitrarily). ■

Define $c_k^*(\boldsymbol{q})$ to be the cost of the shortest route corresponding to class k, i.e., $c_k^*(\boldsymbol{q}) = \min_{p\in\mathcal{T}^k} c_p(\boldsymbol{q})$. Since, $c_k^*(\boldsymbol{q})$ is defined to be the point wise minimum of several linear functions, it is a concave function of $\boldsymbol{q}$ [19]. For several important traffic classes (*e.g.,* unicast, broadcast, anycast) there are standard combinatorial algorithms for efficiently computing $c_k^*(\boldsymbol{q})$ (*e.g.* Weighted Shortest Path, Minimum Weight Spanning Tree etc.).
With an optimal setting of the flow variables $\boldsymbol{f}$ resolved by Lemma (1), Eqn. (23) implies that the computation of the dual objective function (20) reduces to optimizing the traffic admission rates r_k for each class k as follows:

$$r_k^*(\boldsymbol{q}) = \arg\max_{r_k\geq 0} \big(VU_k(r_k) - r_k c_k^*(\boldsymbol{q})\big) \tag{24}$$

Due to strict concavity of the utility functions $U_k(\cdot)$, the optimal solution to the problem (20) is obtained by setting the derivative of the objective with respect to the variable r_k to zero, which yields

$$r_k^*(\boldsymbol{q}) = \bigg(U_k'^{-1}\big(c_k^*(\boldsymbol{q})/V\big)\bigg)^+, \tag{25}$$

where we project r_k on the set of non-negative real numbers due to non-negativity constraints of the rates. Substituting Eqn. (25) into Eqn. (20), we obtain an implicit expression of the dual objective function $D(\boldsymbol{q})$. In the following, we derive an explicit expression of the dual function in terms of the Fenchel conjugate of the utility functions [19].

B. Derivation of the Dual Objective Function

Substituting the value of $r_k^*(\boldsymbol{q})$ into the Lagrangian (19) and noting that $f_p^* = r_k^*(\boldsymbol{q})$ only along the corresponding shortest route and is zero otherwise, we have

$$D(\boldsymbol{q}) = V\sum_k \big(U_k(r_k^*(\boldsymbol{q})) - r_k^*(\boldsymbol{q})\frac{c_k^*(\boldsymbol{q})}{V}\big) + \sum_e q_e c_e. \tag{26}$$

The above expression may also be written in terms of the Fenchel's conjugate [19] of the utility functions. For this, we recall the definition of the Fenchel conjugate $f^\dagger$ of (an extended real-valued) function f:

$$f^\dagger(z) = \sup_{x\in \text{dom}(f)} \big(x^T z - f(x)\big) \tag{27}$$

Now, let the function $U_k^\dagger : \mathbb{R} \to \mathbb{R}$ denote the Fenchel conjugate of the function $-U_k(\cdot)$, which, by our assumption, is a strictly convex function. Thus, from Eqn. (26) we can write

$$D(\boldsymbol{q}) = V\sum_k U_k^\dagger\big(-\frac{c_k^*(\boldsymbol{q})}{V}\big) + \sum_e q_e c_e. \tag{28}$$

In the following, we use Eqn. (28) to derive explicit functional forms of the dual objective functions for two important classes of network utility functions.

1. Logarithmic Utility Functions: Consider the class of Logarithmic utility functions defined as follows:

$$U_k(r_k) = \gamma_k \log(1 + r_k), \quad r_k \geq 0, \tag{29}$$

where γ_k is a positive constant. Among its many attractive properties, the Logarithmic utility functions ensures proportionally fair rate allocations among all participating classes. The Fenchel's conjugate of $-U_k(r_k)$ may be computed as

$$\begin{aligned} U_k^\dagger(z) &= \sup_{x\geq 0} \big(xz + \gamma_k \log(1+x)\big) \\ &= \begin{cases} \gamma_k \log(-\frac{\gamma_k}{z}) - (\gamma_k + z), & \text{if } z < 0 \\ +\infty, & \text{if } z \geq 0 \end{cases} \end{aligned}$$

Hence, the dual function is given as $D(\boldsymbol{q}) =$

$$\begin{cases} V\sum_k \big(\gamma_k \log(\frac{\gamma_k V}{c_k^*(\boldsymbol{q})}) - \gamma_k + \frac{c_k^*(\boldsymbol{q})}{V}\big) + \sum_e q_e c_e, & \text{if } c_k^*(\boldsymbol{q}) > 0 \\ +\infty, & \text{if } c_k^*(\boldsymbol{q}) \leq 0, \end{cases}$$

2. α-fair Utility Functions: Next, we consider the α-fair utility functions defined as follows:

$$U_k(r_k) = \gamma_k \frac{r_k^{1-\alpha}}{1-\alpha}, \quad r_k \geq 0, \tag{30}$$

where $\gamma_k > 0$ and $0 < \alpha < 1$ are positive parameters.
The Fenchel's conjugate of $-U_k(r_k)$ may be computed as

$$\begin{aligned} U_k^\dagger(z) &= \sup_{x\geq 0} \bigg(xz + \gamma_k \frac{x^{1-\alpha}}{1-\alpha}\bigg) \\ &= \begin{cases} \frac{\alpha}{1-\alpha}\gamma_k^{1/\alpha}(-z)^{1-1/\alpha}, & \text{if } z < 0 \\ +\infty, & \text{if } z \geq 0 \end{cases} \end{aligned}$$

Hence, the dual function $D(\boldsymbol{q})$ is given as

$$D(\boldsymbol{q}) = \begin{cases} V\frac{\alpha}{1-\alpha}\sum_k \gamma_k^{1/\alpha}\big(\frac{c_k^*(\boldsymbol{q})}{V}\big)^{1-1/\alpha} + \sum_e q_e c_e, & \text{if } c_k^*(\boldsymbol{q}) > 0 \\ +\infty, & \text{if } c_k^*(\boldsymbol{q}) \leq 0, \end{cases}$$

C. Subgradient Method and its Equivalence with UMW+

Since the dual objective $D(\boldsymbol{q})$, as given in Eqn. (20), is a point wise maximum of linear functions, it is convex [19]. Moreover, the objective $D(\boldsymbol{q})$, as seen from Eqn. (28), is *non-differentiable*, as the shortest path cost $c_k^*(\boldsymbol{q})$ is not a differentiable function of $\boldsymbol{q}$, in general. Hence, we use a first order method suitable for non-smooth objectives, known as the *Subgradient Descent* [19], to solve the dual problem (22). A subgradient $\boldsymbol{g} \in \partial D(\boldsymbol{q})$ of the dual function $D(\boldsymbol{q})$ may be computed directly from the capacity constraint of the primal problem (Eqn. (17)) as follows:

$$g_e(\boldsymbol{q}) = c_e - A_e(\boldsymbol{q}), \quad \forall e \in E, \tag{31}$$

where $A_e(\boldsymbol{q}) \equiv \sum_{p:e\in p} f_p^*(\boldsymbol{q}) = \sum_k r_k^*(\boldsymbol{q})\mathbb{1}(e \in p_k^*(\boldsymbol{q}))$. Finally, we solve the dual problem $\mathcal{P}^*$ by the dual subgradient method with constant step-size equal to $\theta > 0$. This yields the following iteration for the dual variables $\boldsymbol{q}$:

$$q_e(t+1) = (q_e(t) + \theta(A_e(\boldsymbol{q}) - c_e))^+, \quad \forall e \in E. \tag{32}$$

For the step-size $\theta = 1$, the above iteration *corresponds exactly* to the virtual queue dynamics. More precisely, Eqn. (8) corresponds to Lemma 1, which is concerned with routing along the shortest route in the weighted graph; Eqn. (10) corresponds to Eqn. (25), which corresponds to the packet admission control, and the virtual queue dynamics in Eqn. (4) corresponds to the dual subgradient update of Eqn. (32). Moreover, for any constant step-size $\theta > 0$, at any step t, the virtual queues $\tilde{\boldsymbol{Q}}(t)$ under UMW+ are exactly equal to the $1/\theta$-scaled versions of the corresponding dual variables $\boldsymbol{q}(t)$ under the subgradient descent iterations.

Convergence of the Subgradient Descent: To establish the convergence of the subgradient iterations, it is useful to add an additional constraint in the primal problem $\mathcal{P}$, without changing the optimal solution. For this, observe that, for any feasible solution to $\mathcal{P}$, from the capacity constraint (17), we have

$$\sum_k r_k \leq \sum_e c_e.$$

With this additional constraint, we conveniently restrict the range of each of the admission control variables r_k, in the sub-problems (24), to $[0, \sum_e c_e]$. This upper-bounds the rate of flows crossing the edge e to $A_e(\boldsymbol{q}) \leq |\mathcal{K}| \sum_e c_e \stackrel{\text{def}}{=} \bar{A}_{\max}$. Hence, the norms of the subgradients $\boldsymbol{g}(t)$ (from Eqn. (31)) are *uniformly* bounded by

$$||\boldsymbol{g}(t)||_2^2 \leq \sum_e (c_e^2 + A_e^2(\boldsymbol{q}(t))) \leq m(c_{\max}^2 + m\bar{A}_{\max}^2).$$

We have the following convergence result:

Theorem 3 (Convergence within a Neighborhood)**.** *Under the subgradient iterations* (32)*, we have*

$$VU^* + \frac{\theta b^2}{2} \geq \limsup_{t\to\infty} D(\boldsymbol{q}(t)) \geq \liminf_{t\to\infty} D(\boldsymbol{q}(t)) \geq VU^*,$$

where $b^2 \equiv m(c_{\max}^2 + m|\mathcal{K}|^2(\sum_e c_e)^2)$.

Theorem 3 follows from an application of Proposition 2.2.2 of [20]. This shows that the utility achieved by the dual algorithm is within an additive gap of $O(1/V)$ from U^*.

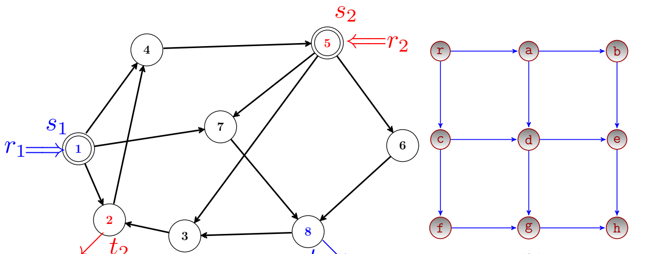

Fig. 1: Network Topologies used for (a) Unicast and (b) Broadcast Simulations

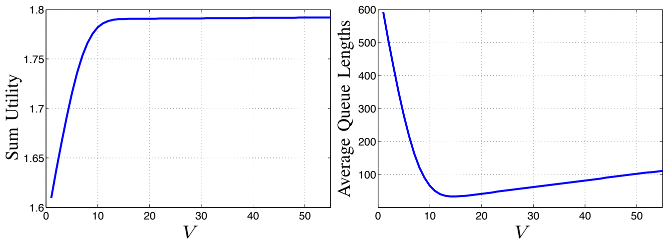

Fig. 2: Performance of the UMW+ policy for the Utility Maximization problem (NUM).

VI. Numerical Simulations

In this Section, we provide simulation results to explore the performance of the UMW+ policy in diverse network and traffic settings.

A. Unicast Traffic in a Wired Network

To begin with, we consider the same wired network topology as in [1], with two unicast sources s_1, s_2 and destinations t_1, t_2, shown in Figure 1 (a). The capacity of each directed link is taken to be one packet per slot. From the network topology, it can be easily seen that there are two $s_1 \to t_1$ paths (*e.g.*, $1 \to 4 \to 5 \to 6 \to 8$ and $1 \to 7 \to 8$) and one $s_2 \to t_2$ path (*e.g.*, $5 \to 3 \to 2$), which are *mutually edge disjoint*. Moreover, we also have $\textsc{cut}(\{1,2,3,4,5,6,7\},\{8\}) = 2$ and $\textsc{cut}(\{4,5,6,7,8\},\{1,2\}) = 1$. This implies that the optimal solution to the NUM problem is attained at $r_1 = 2, r_2 = 1$ for any non-decreasing utility function.

Next, we run the proposed UMW+ policy with logarithmic utility functions $U_1(r_1) = \ln(1+r_1), U_2(r_2) = \ln(1+r_2)$. The theoretical optimal utility value, in this case, is easily computed to be $U^* = \ln(3) + \ln(2) \approx 1.79$. Figure 2 (a) shows the achieved utility by the UMW+ policy with the variation of the V parameter. Figure 2 (b) shows the variation of the total queue length as a function of V. These figures clearly demonstrate the the utility-optimality of the UMW+ policy.

Figure 3 shows the temporal dynamics of the subgradient algorithm for different V parameters, and two different step-sizes, $\theta = 1$ and $\theta = 0.1$. Note that, any solution to the dual problem gives an upper bound of the primal objective (weak duality). The optimal solution of the dual minimization problem has zero duality gap with the optimal primal solution.

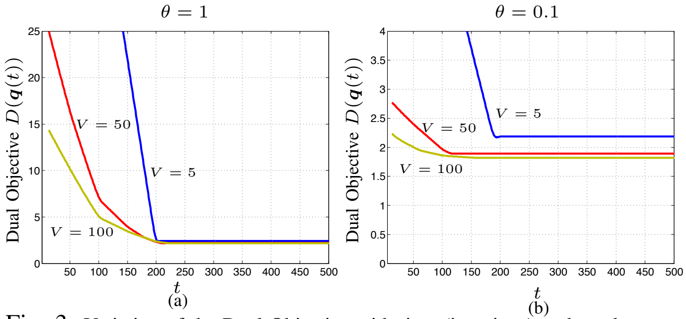

Fig. 3: Variation of the Dual Objective with time (iterations) under subgradient descent for step-sizes $\theta = 1$, and $\theta = 0.1$.

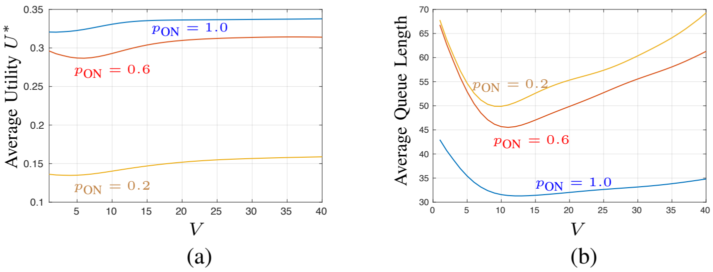

Fig. 4: Variation of time-averaged Utility and total physical queue lengths with the parameter V under the UMW+ policy in the broadcast setting for the time-varying grid network in Fig 1 (b).

It is evident from the plots in Figure 3 that the speed of convergence increases and the optimality gap decreases with the increase of the parameter V. This observation is consistent with the statement of Theorem 3, which states that the solution obtained by the dual algorithm lies within an additive gap of $O(1/V)$ from the optimal utility U^*.

Broadcast Traffic in a Time-varying Wireless Network

Next, we simulate the UMW+ policy for a time-varying wireless network, shown in Figure 1 (b). Each wireless link is ON i.i.d. at every slot with probability p_{ON}. Link activations are limited by primary interference constraints. We use the same logarithmic utility function as before. Under the proposed UMW+ policy, variation of the average utility and average queue lengths with the parameter V is shown in Figure 4 (a) and 4 (b), for three different values of p_{ON}. As expected, with better average channel conditions (i.e., higher values of p_{ON}) higher utility is achieved with smaller in-network queue lengths.

VII. Conclusion

In this paper, we have proposed the first network control policy, called UMW+, to solve the Utility Maximization Problem with multiple types of concurrent traffic, including unicast, broadcast, multicast, and anycast. The proposed policy effectively exploits the novel idea of *precedence-relaxation* of a multi-hop network. We relate the UMW+ policy to the subgradient iterations of an associated dual problem. The dual objective function of the associated static NUM problem has been characterized in terms of the Fenchel conjugates of the associated utility function. Finally, illustrative simulation results have been provided for both wired and wireless networks under different input traffic settings.

References

[1] A. Sinha and E. Modiano, "Optimal control for generalized network-flow problems," *IEEE/ACM Transactions on Networking*, vol. 26, no. 1, pp. 506–519, Feb 2018.

[2] F. P. Kelly, A. K. Maulloo, and D. K. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research society*, vol. 49, no. 3, pp. 237–252, 1998.

[3] J. W. Lee, R. R. Mazumdar, and N. B. Shroff, "Downlink power allocation for multi-class cdma wireless networks," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 2002, pp. 1480–1489.

[4] S. H. Low, "A duality model of tcp and queue management algorithms," *IEEE/ACM Transactions On Networking*, vol. 11, no. 4, pp. 525–536, 2003.

[5] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, 2007.

[6] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 2, pp. 396–409, 2008.

[7] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on Selected areas in Communications*, vol. 24, no. 8, pp. 1452–1463, 2006.

[8] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," *IEEE/ACM Transactions on Networking (TON)*, vol. 15, no. 6, pp. 1333–1344, 2007.

[9] V. Valls and D. J. Leith, "On the relationship between queues and multipliers," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*. IEEE, 2014, pp. 250–255.

[10] A. Sinha and E. Modiano, "Throughput-optimal broadcast in wireless networks with point-to-multipoint transmissions," in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. Mobihoc '17. New York, NY, USA: ACM, 2017, pp. 3:1–3:10.

[11] J. Zhang, A. Sinha, J. Llorca, A. Tulino, and E. Modiano, "Optimal Control of Distributed Computing Networks with Mixed-Cast Traffic Flows," in *Computer Communications (INFOCOM), 2018 IEEE Conference on*. IEEE, 2018.

[12] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[13] A. Sinha, P. Mani, J. Liu, A. Flavel, and D. A. Maltz, "Distributed load management algorithms in anycast-based cdns," *Comput. Netw.*, vol. 115, no. C, pp. 1–15, Mar. 2017. [Online]. Available: https://doi.org/10.1016/j.comnet.2017.01.016

[14] J.-H. Hoepman, "Simple distributed weighted matchings," *arXiv preprint cs/0410047*, 2004.

[15] A. Sinha and E. Modiano, "Network utility maximization with heterogeneous traffic flows," Technical Report. [Online]. Available: https://arxiv.org/abs/1803.00081

[16] L. Kleinrock, *Queueing system volume 1: Theory*. John Wiley&Sons, New York, 1975.

[17] D. Gamarnik, "Stability of adaptive and non-adaptive packet routing policies in adversarial queueing networks," in *Proceedings of the thirty-first annual ACM symposium on Theory of computing*. ACM, 1999.

[18] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *Automatic Control, IEEE Transactions on*, vol. 37, no. 12, pp. 1936–1948, 1992.

[19] D. P. Bertsekas, A. Nedich, A. E. Ozdaglar *et al.*, *Convex analysis and optimization*. Athena Scientific, 2003.

[20] D. P. Bertsekas and A. Scientific, *Convex optimization algorithms*. Athena Scientific Belmont, 2015.