Resource Allocation: Realizing Mean-Variability-Fairness Tradeoffs

Vinay Joseph, Gustavo de Veciana, *Fellow, IEEE*, and Ari Arapostathis, *Fellow, IEEE*

***Abstract*—Network utility maximization (NUM) is a key conceptual framework to study reward allocation amongst a collection of users/entities in disciplines as diverse as economics, law and engineering. However when the available resources and/or users' utilities vary over time, reward allocations will tend to vary, which in turn may have a detrimental impact on the users' overall satisfaction or quality of experience. This paper introduces a generalization of the NUM framework which incorporates the detrimental impact of temporal variability in a user's allocated rewards. It explicitly incorporates tradeoffs amongst the mean and variability in users' reward allocations, as well as fairness across users. We propose a simple online algorithm to realize these tradeoffs, which, under stationary ergodic assumptions, is shown to be asymptotically optimal, i.e., achieves a long term performance equal to that of an offline algorithm with knowledge of the future variability in the system. This substantially extends work on NUM to an interesting class of relevant problems where users/entities are sensitive to temporal variability in their service or allocated rewards.**

***Index Terms*—Network utility maximization (NUM).**

I. Introduction

NETWORK utility maximization (NUM) is a key conceptual framework to study (fair) reward allocation among a collection of users/entities across disciplines as diverse as economics, law and engineering. For example, [25] introduces NUM for realizing fair allocations of a *fixed* amount of water c to N farms. The amount of water w_i allocated to the ith farm is a resource which yields a reward $r_i = f_i(w_i)$ to the ith farm. Here, f_i is a concave function mapping allocated water (resource) to yield (reward), and these can differ across farms. The allocation maximizing $\sum_{1\le i\le N} r_i$ is a reward (utility) maximizing solution to the problem. Fairness can be imposed on the allocation by changing the objective of the problem to $\sum_{1\le i\le N} U(r_i)$ for an appropriately chosen concave function U. Now, suppose that we have to make allocation decisions periodically to respond to time varying water availability $(c_t)_{t\in\mathbb{N}}$ and utility functions $(f_{i,t})_t$. Then, subject to the time varying constraints, one could obtain a resource allocation scheme which is fair in the delivery of time average rewards $\overline{\mathbf{r}} = (\overline{r}_i)_{1\le i\le N}$ by optimizing (see, e.g., [16], [30])

$$\sum_{1\le i\le N} U(\overline{r}_i). \tag{1}$$

In network engineering, the NUM framework has served as a particularly insightful setting to study (reverse engineer) how the Internet's congestion control protocols allocate bandwidth, how to devise schedulers for wireless systems with time varying channel capacities, and also motivated the development of distributed mechanisms to maximize network utility in diverse settings including communication networks and the smart grid, while incorporating new relevant constraints, on energy, power, storage, power control, stability, etc. (for, e.g., see [14], [25], [30]).

When the available resources/rewards and/or users' utilities vary over time, reward allocations amongst users will tend to vary, which in turn may have a detrimental impact on the users' utility or perceived service quality. In fact, temporal variability in farm water availability can have a negative impact on crop yield (see [28]). This motivates modifications of formulations with objectives such as the one in (1) to account for this impact.

Indeed temporal variability in utility, service, rewards or associated prices is particularly problematic when humans are the eventual recipients of the allocations. Humans typically view temporal variability negatively, as a sign of an unreliable service, network or market instability. Broadly speaking, temporal variability, when viewed through human's cognitive and behavioral responses, leads to a degraded Quality of Experience (QoE). This in turn can lead users to make decisions, e.g., change provider, act upon perceived market instabilities, etc., which can have serious implications on businesses and engineered systems, or economic markets. For problems involving resource allocation in networks, [5] argues that predictable or consistent service is essential and even points out that it may be appropriate to intentionally lower the quality delivered to the user if that level is sustainable.

For a user viewing a video stream, variations in video quality over time have a detrimental impact on the user's QoE, see e.g., [15], [23], [33]. Indeed [33] suggested that variations in quality can result in a QoE that is worse than that of a constant quality video with lower average quality. Furthermore, [33] proposed a metric for QoE given below which penalizes standard deviation of quality over time

$$\text{Mean Quality} - \kappa\sqrt{\text{Temporal Variance in Quality}}$$

Manuscript received October 20, 2012; revised September 11, 2013 and April 4, 2014; accepted May 11, 2014. Date of publication June 25, 2014; date of current version December 22, 2014. This research was supported in part by Intel and Cisco under the VAWN program, and by the NSF under Grant CNS-0917067. The research of Ari Arapostathis was supported in part by the Office of Naval Research through the Electric Ship Research and Development Consortium. Recommended by Associate Editor L. H. Lee.

The authors are with The University of Texas at Austin, Austin, TX 78712 USA (e-mail: vinayjoseph@mail.utexas.edu; gustavo@ece.utexas.edu; ari@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2014.2332731

where κ is an appropriately chosen positive constant. References [9] and [32] argue that less variability in the service processes can improve customer satisfaction by studying data for large retail banks and major airlines respectively. Aversion towards temporal variability is not just restricted to human behavior, for instance, see [22] for a discussion of the impact of temporal variability in nectar reward on foraging behavior of bees. Also, variability in resource allocation in networks can lead to burstiness which can degrade network performance (see [7], [24]). These examples illustrate the need for extending the NUM framework to incorporate the impact of variability.

This paper introduces a generalized NUM framework which explicitly incorporates the detrimental impact of temporal variability in a user's allocated rewards. We use the term rewards as a proxy for the resulting utility of, or any other quantity associated with, allocations to users/entities in a system. Our goal is to explicitly tackle the task of incorporating tradeoffs amongst the mean and variability in users' rewards. Thus, for example, in a variance-sensitive NUM setting, it may make sense to reduce a user's mean reward so as to reduce his/her variability. As will be discussed in the sequel, there are many ways in which temporal variations can be accounted for, and which, in fact, present distinct technical challenges. In this paper, we shall take a simple elegant approach to the problem which serves to address systems where tradeoffs amongst the mean and variability over time need to be made rather than systems where the desired mean (or target) is known (as in minimum variance control, see [2]), or where the issue at hand is minimization of the variance of a cumulative reward at the end of a given (e.g., investment) period.

To better describe the characteristics of the problem we introduce some preliminary notation. We shall consider a network shared by a set $\mathcal{N}$ of users (or other entities) where $N := |\mathcal{N}|$ denotes the number of users in the system. Throughout the paper, we distinguish between random variables (and random functions discussed later) and their realizations by using upper case letters for the former and lower case for the latter. Let $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{R}_+$ denote the sets of positive integers, real numbers and nonnegative real numbers respectively. We use bold letters to denote vectors, e.g., $\mathbf{a} = (a_i)_{i\in\mathcal{N}}$. Given a collection of T objects $(b(t))_{1\le t\le T}$ or a sequence $(b(t))_{t\in\mathbb{N}}$, we let $(b)_{1:T}$ denote the finite length sequence $(b(t))_{1\le t\le T}$ (in the space associated with the objects of the sequence). For example, consider a sequence $(\mathbf{b}(t))_{t\in\mathbb{N}}$ of vectors in $\mathbb{R}^N$. Then $(\mathbf{b})_{1:T}$ denotes the T length sequence containing the first T vectors of the sequence $(\mathbf{b}(t))_{t\in\mathbb{N}}$, and $(b_i)_{1:T}$ denotes the sequence containing the ith component of the first T vectors. For any function U on $\mathbb{R}$, let U' denote its derivative.

Definition 1: For any (infinite length) sequence of real numbers $(a(t))_{t\in\mathbb{N}}$, let

$$m^T(a) := \frac{1}{T}\sum_{t=1}^{T} a(t)$$

$$\mathrm{Var}^T(a) := \frac{1}{T}\sum_{t=1}^{T}\left(a(t) - m^T(a)\right)^2$$

$$e_i^T(a) := m^T(a) - U_i^V\left(\mathrm{Var}^T(a)\right)$$

i.e., $m^T(a)$ and $\mathrm{Var}^T(a)$ denote empirical mean and variance. Note that the argument a used in the functions $m^T(a)$, $\mathrm{Var}^T(a)$ and $e_i^T(a)$ stands for the associated sequence $(a(t))_{t\in\mathbb{N}}$. We will also (abusing notation) use the above operators on any finite length sequence $(a)_{1:T} \in \mathbb{R}^T$ of real numbers.

Let $r_i(t)$ represent the reward allocated to user i at time t. Then $\mathbf{r}(t) = (r_i(t))_{i\in\mathcal{N}}$ is the vector of rewards to users $\mathcal{N}$ at time t, and $(\mathbf{r})_{1:T}$ represents sequence of vector rewards allocated over time slots $t = 1, \ldots, T$. We assume that reward allocations are subject to time varying network constraints

$$c_t\left(\mathbf{r}(t)\right) \le 0 \quad \text{for} \quad t = 1, \ldots, T$$

where each $c_t : \mathbb{R}^N \to \mathbb{R}$ is a convex function, thus implicitly defining a convex set of feasible reward allocations. To formally capture the impact of the time-varying rewards on users' QoE, let $(U_i^E, U_i^V)_{i\in\mathcal{N}}$ be real valued functions on $\mathbb{R}$, and consider the following *offline* convex optimization problem OPT(*T*):

$$\max_{(\mathbf{r})_{1:T}} \sum_{i\in\mathcal{N}} U_i^E\left(\overbrace{\underbrace{m^T(r_i)}_{\text{Mean Reward}} - \underbrace{U_i^V\left(\mathrm{Var}^T(r_i)\right)}_{\text{Penalty for Variability}}}^{\text{User } i\text{'s QoE}}\right)$$

$$\text{subject to} \quad c_t\left(\mathbf{r}(t)\right) \le 0, \mathbf{r}(t) \ge \mathbf{0}\ \forall t \in \{1, \ldots, T\}.$$

We refer to OPT(*T*) as an offline optimization because time-varying time constraints $(c_t)_{1:T}$ are assumed to be known. Here, $(U_i^E, U_i^V)_{i\in\mathcal{N}}$ are increasing functions such that the above optimization problem is convex. For user i, the argument of the function U_i^E is our proxy for the user's QoE. Thus, the desired fairness in the allocation of QoE across the users can be imposed by appropriately choosing $(U_i^E)_{i\in\mathcal{N}}$. Note that the first term $m^T(r_i)$ in user i's QoE is the user's mean reward allocation, whereas the presence of the empirical variance function $\mathrm{Var}^T(r_i)$ in the second term penalizes temporal variability in a reward allocation. Further, flexibility in picking $(U_i^V)_{i\in\mathcal{N}}$ allows for several different ways to penalize such variability. Indeed, one can in principle have a variability penalty that is convex or concave in variance. Hence, the formulation $\mathrm{OPT}(T)$ allows us to realize tradeoffs among mean, fairness and variability associated with the reward allocation by appropriately choosing the functions $(U_i^E, U_i^V)_{i\in\mathcal{N}}$.

A. Main Contributions

The main contribution of this paper is the development of an *online* algorithm, Adaptive Variability-aware Reward allocation (AVR), which asymptotically solves OPT(*T*). The algorithm requires almost no statistical information about the system, and its characteristics are as follows:

(i) in each time slot, c_t is revealed and AVR, using parameters $\mathbf{m}(t), \mathbf{v}(t) \in \mathbb{R}^N$, *greedily* allocates rewards by solving the optimization problem OPT-ONLINE given below:

$$\max_{\mathbf{r}\in\mathbb{R}^N} \sum_{i\in\mathcal{N}} \left(U_i^E\right)'\left(e_i(t)\right)\left(r_i - \left(U_i^V\right)'\left(v_i(t)\right)\left(r_i - m_i(t)\right)^2\right)$$

$$\text{subject to } c_t(\mathbf{r}) \le 0, \quad \mathbf{r} \ge 0$$

where $e_i(t) = m_i(t) - U_i^V(v_i(t))$ for each $i \in \mathcal{N}$ is an estimate of the user's QoE based on estimated means and variances $\mathbf{m}(t)$ and $\mathbf{v}(t)$; and,

(ii) it updates (vector) parameters $\mathbf{m}(t)$ and $\mathbf{v}(t)$ to keep track of the mean and variance of the reward allocations under AVR.

Under stationary ergodic assumptions for time-varying constraints, we show that our *online* algorithm AVR is asymptotically optimal, i.e., achieves a performance equal to that of the *offline* optimization OPT(T) introduced earlier as $T \to \infty$. This is a strong optimality result, which at first sight may be surprising due to the variability penalty on rewards and the time varying nature of the constraints $(c_t)_{t\in\mathbb{N}}$. The key idea is to keep online estimates for the relevant quantities associated with users' reward allocations, e.g., the mean and variance which over time are shown to converge. This in turn eventually enables our greedy online policy to produce reward allocations corresponding to the optimal stationary policy. Proving this result is somewhat challenging as it requires showing that the estimates based on reward allocations produced by our online policy, AVR, (which itself depends on the estimated quantities), will converge to the desired values. To our knowledge this is the first attempt to generalize the NUM framework in this direction. We contrast our problem formulation and approach to past work in addressing 'variability' minimization, risk-sensitive control and other MDP based frameworks in the next subsection.

B. Related Work

NUM is a well studied approach used for reward allocation amongst a collection of users/entities. The work in [25] provides a network-centric overview of NUM. All the work on NUM including several major extensions (for, e.g., [14], [21], [29], [30] etc.) has ignored the impact of variability in reward allocation. Our work [12] is to our knowledge the first to tackle NUM incorporating the impact of variability explicitly. In particular, we addressed a special case of the problem studied in this paper that only allows for linear functions $(U_i^E, U_i^V)_{i\in\mathcal{N}}$, and an asymptotically optimal online reward allocation algorithm for a wireless network supporting video streaming users is proposed. The algorithm proposed and analyzed in this paper is a generalization of gradient based algorithms studied in [1], [16], and [30]. Our approach for proving asymptotic optimality generalizes those in [13] and [30]. In [30], the focus is on objectives such as (1), but does not allow for the addition of penalty terms on temporal variance in the objective. By contrast with this paper, the approaches in [12] and [13] rely on the use of results on sensitivity analysis of optimization problems, and only allows for linear $(U_i^E)_{i\in\mathcal{N}}$ and concave $(U_i^V)_{i\in\mathcal{N}}$.

Adding a temporal variance term in the cost takes the objective out of the basic dynamic programming setting (even when $(U_i^E, U_i^V)_{i\in\mathcal{N}}$ are linear) as the overall cost is not decomposable over time, i.e., can not be written as a sum of costs each depending only on the allocation at that time- this is what makes sensitivity to variability challenging. For risk sensitive decision making, MDP based approaches aimed at realizing optimal tradeoffs between mean and temporal variance in reward/cost were proposed in [8] and [26]. While they consider a more general setting than ours where actions can even affect future feasible reward allocations, e.g., may affect the process $(C_t)_{t\in\mathbb{N}}$ itself, the approaches proposed in these works suffer from the curse of dimensionality as they require solving large optimization problems. For instance, the work of [8] involves solving a quadratic program in the (typically large) space of state-action pairs. Note that these works on risk sensitive decision making are different from those focusing on the variance of the *cumulative* cost/reward such as the one in [19].

Variability or perceived variability can be measured in many different ways, and temporal variance considered in this paper is one of them. One could also 'reduce variability' using a minimum variance controller (see [2]) where we have certain target reward values fixed ahead of time and big fluctuations from these targets are undesirable. Note however that in using this approach, we have to fix our targets ahead of time, and thus lose the ability to realize tradeoffs between the mean and variability in reward allocation. One could also measure variability using switching costs like in [18], which consider the problem of achieving tradeoffs between average cost and time average switching cost associated with data center operation, and proposes algorithms with good performance guarantees for adversarial scenarios. The decision regarding how to penalize variability is ultimately dependent on the application setting under consideration.

C. Organization of the Paper

Section II introduces the system model and assumptions. Section III presents and studies the offline formulation for optimal variance sensitive joint reward allocation OPT(T). Section IV formally introduces our online algorithm AVR and presents our key convergence result which is used to prove asymptotic optimality of AVR. Section V is devoted to the proof of AVR's convergence and Section VI presents simulation results exhibiting additional performance characteristics of AVR. We conclude the paper with Section VII. Proofs for some of the results have been relegated to the Appendices to make the paper more readable.

II. System Model

We consider a slotted system where time slots are indexed by $t \in \mathbb{N}$, and the system serves a fixed set of users $\mathcal{N}$ and let $N := |\mathcal{N}|$.

We assume that rewards are allocated subject to time varying constraints. The reward allocation $\mathbf{r}(t) \in \mathbb{R}_+^N$ in time slot t is constrained to satisfy the following inequality:

$$c_t\left(\mathbf{r}(t)\right) \leq 0$$

where c_t denotes the realization of a randomly selected function C_t from a finite set $\mathcal{C}$ of real valued maps on $\mathbb{R}_+^N$. We model the reward constraints $(C_t)_{t\in\mathbb{N}}$ as a random process where each C_t can be viewed as a random function, i.e., a random index

associated with the function (which is selected from a finite set $\mathcal{C}$). We make the following assumptions on these constraints:

Assumptions C1–C3 (Time varying constraints on rewards)

C.1 $(C_t)_{t\in\mathbb{N}}$ is a stationary ergodic process of functions selected from a finite set $\mathcal{C}$.

C.2 The feasible region for each constraint is bounded: there is a constant $0 < r_{\max} < \infty$ such that for any $c \in \mathcal{C}$ and $\mathbf{r} \in \mathbb{R}_+^N$ satisfying $c(\mathbf{r}) \leq 0$, we have $r_i \leq r_{\max}$ for each $i \in \mathcal{N}$.[1]

C.3 Each function $c \in \mathcal{C}$ is convex and differentiable on an open set containing $[0, r_{\max}]^N$ with $c(\mathbf{0}) \leq 0$ and

$$\min_{\mathbf{r}\in[0,r_{\max}]^N} c(\mathbf{r}) < 0. \tag{2}$$

As indicated in Assumption C.1, we model the evolution of the reward constraints as a stationary ergodic process. Hence, time averages associated with the constraints will converge to their respective statistical averages, and the distribution of the random vector $(C_{t_1+s}, C_{t_2+s}, \ldots, C_{t_n+s})$ for any choice of indices $t_1, \ldots, t_n$ does not depend on the shift s, thus the marginal distribution of C_t does not depend on time. We denote the marginal distribution of this process by $(\pi(c))_{c\in\mathcal{C}}$ and let C^π denote a random constraint with this distribution. This model captures a fairly general class of constraints, including, for example, time-varying capacity constraints associated with bandwidth allocation in wireless networks. Our results also hold when $(C_t)_{t\in\mathbb{N}}$ is an asymptotically mean stationary process (see [11] for reference) of functions selected from a finite set $\mathcal{C}$ and this is discussed in Section IV. If condition C.2 holds, then we can upper bound any feasible allocation under any constraint in $\mathcal{C}$ using $r_{\max}\mathbf{1}_N$ where $\mathbf{1}_N$ is the N length vector with each component equal to one. Condition C.3 ensures that the feasible sets are convex, and the differentiability requirement simplifies the exposition. The remaining requirements in C.3 are useful in studying the optimization problem $\mathrm{OPT}(T)$.

Next we introduce the assumptions on the functions $(U_i^V)_{i\in\mathcal{N}}$ associated with the variability penalties.

Assumptions U.V: (Variability penalty) Let $v_{\max} := r_{\max}^2$.

U.V.1: For each $i \in \mathcal{N}$, U_i^V is well defined and differentiable on an open set containing $[0, v_{\max}]$ satisfying $\min_{v\in[0,v_{\max}]}(U_i^V)'(v) > 0$, and $(U_i^V)'(\cdot)$ is Lipschitz continuous.

U.V.2: For each $i \in \mathcal{N}$ and any $z_1, z_2 \in [-\sqrt{v_{\max}}, \sqrt{v_{\max}}]$ with $z_1 \neq z_2$, and $\alpha \in (0,1)$ with $\bar{\alpha} = 1 - \alpha$, we have

$$U_i^V\left((\alpha z_1 + \bar{\alpha} z_2)^2\right) < \alpha U_i^V\left(z_1^2\right) + \bar{\alpha} U_i^V\left(z_2^2\right). \tag{3}$$

The assumptions concerning the Lipschitz continuity of derivatives made in Assumptions U.V.1 and U.E (see below) are made to simplify the exposition, and could be relaxed (see Section V-B). Note that any non-decreasing (not necessarily strictly) convex function satisfies (3), but the condition is weaker than a convexity requirement. For instance, using triangle inequality, one can show that $U_i^V(v_i) = \sqrt{v_i + \delta}$ for $\delta > 0$ satisfies all the conditions described above for any $v_{\max}$.[2] This function is not convex but is useful as it transforms variance to approximately the standard deviation for small $\delta > 0$, and thus allowing QoE metrics such as those proposed in [33] (discussed in Section I). We will later see that our algorithm (Section I-A) can be simplified if any of the functions U_i^V are linear. Hence, we define the following subsets of $\mathcal{N}$:

$$\mathcal{N}_l := \left\{i \in \mathcal{N} : U_i^V \text{ is linear}\right\}$$
$$\mathcal{N}_n := \left\{i \in \mathcal{N} : U_i^V \text{ is not linear}\right\}.$$

Next we discuss assumptions on the functions $(U_i^E)_{i\in\mathcal{N}}$ used to impose fairness associated with the QoE across users. Recall that our proxy for the QoE for user i is $e_i(t) = m_i(t) - U_i^V(v_i(t))$ and, let

$$e_{\min,i} := -U_i^V(v_{\max}) \text{ and } e_{\max,i} := r_{\max} - U_i^V(0).$$

Assumption U.E: (Fairness in QoE)

U.E: For each $i \in \mathcal{N}$, U_i^E is concave and differentiable on an open set containing $[e_{\min,i}, e_{\max,i}]$ with $(U_i^E)'(e_{\max,i}) > 0$, and $(U_i^E)'(\cdot)$ is Lipschitz continuous.

Note that concavity and the condition that $(U_i^E)'(e_{\max,i}) > 0$ ensure that $(U_i^E)'$ is strictly positive on $[e_{\min,i}, e_{\max,i}]$. For each $i \in \mathcal{N}$, although U_i^E has to be defined over an open set containing $[e_{\min,i}, e_{\max,i}]$, only the definition of the function over $[-U_i^V(0), e_{\max,i}]$ affects the optimization. This is because we can achieve this value of QoE for each user just by allocating zero resources to each user in each time slot. Thus, for example, we can choose any function from the following class of strictly concave increasing functions parametrized by $\alpha \in (0, \infty)$ [20]:

$$U_\alpha(e) = \begin{cases} \log(e) & \text{if } \alpha = 1 \\ (1-\alpha)^{-1} e^{1-\alpha} & \text{otherwise} \end{cases} \tag{4}$$

and can satisfy U.E by making minor modifications to the function. For instance, we can use the following modification $U^{E,log}$ of the log function for any (small) $\delta > 0$: $U^{E,log}(e) = \log(e - e_{\min,i} + \delta),\ e \in [e_{\min,i}, e_{\max,i}]$. The above class of functions are commonly used to enforce fairness specifically to achieve reward allocations that are α-fair (see [25]).

Good choices of $(U_i^V)_{i\in\mathcal{N}}$ and $(U_i^E)_{i\in\mathcal{N}}$ will depend on the problem setting. A good choice for $(U_i^V)_{i\in\mathcal{N}}$ should be driven by an understanding of the impact of temporal variability on a user's QoE, which might in turn be based on experimental data. For instance, a choice of $U_i^V(v_i) = \sqrt{v_i + \delta}$ is proposed for video adaptation in [33]. The choice of $(U_i^E)_{i\in\mathcal{N}}$ is driven by the degree of fairness in the allocation of QoE across users, e.g., max-min, proportional fairness etc. A larger α corresponds to a

[1] We could allow the constant $r_{\max}$ to be user dependent. But, we avoid this for notational simplicity.

[2] Note that we need $\delta > 0$ otherwise $U_i^V(v_i) = \sqrt{v_i}$ violates U.V.1.

more fair allocation which eventually becomes max-min fair as α goes to infinity.

Applicability of the Model: We close this section by illustrating the wide scope of the framework discussed above by describing examples of scenarios that fit it nicely. They illustrate the freedom provided by the framework for modeling temporal variability in both the available rewards and the sensitivity of the users' reward/utility to their reward allocations, as well as fairness across users' QoE. The presence of time-varying constraints $c_t(\mathbf{r}) \leq 0$ allows us to apply the model to several interesting settings. In particular, we discuss three wireless network settings and show that the framework can handle problems involving time-varying exogenous loads and time-varying utility functions.

1) Time-Varying Capacity Constraints: We start by discussing the case where the rewards in a time slot is the rate allocated to the users, and users dislike variability in their allocations. Let $\mathcal{P}$ denote a finite (but arbitrarily large) set of positive vectors where each vector corresponds to the peak transmission rates achievable to the set of users in a given time slot. Let $\mathcal{C} = \{c_{\mathbf{p}} : c_{\mathbf{p}}(\mathbf{r}) = \sum_{i\in\mathcal{N}}(r_i/p_i) - 1, \mathbf{p} \in \mathcal{P}\}$. Here, for any allocation $\mathbf{r}$, r_i/p_i is the fraction of time the wireless system needs to serve user i in time slot t in order to deliver data at the rate of r_i when the user has peak transmission rate p_i. Thus, the constraint $c_{\mathbf{p}}(\mathbf{r}) \leq 0$ can be seen as a scheduling constraint that corresponds to the requirement that the sum of the fractions of time that different users are served in a time slot should be less than or equal to one.

2) Time-Varying Exogenous Constraints: We can further introduce time varying exogenous constraints on the wireless system by appropriately defining the set $\mathcal{C}$. For instance, consider a base station in a cellular network that supports users who dislike variability in rate allocation. But, while allocating rates to these users, we may also need to account for the time-varying rate requirements of the voice traffic handled by the base station. We can model this by defining

$$\mathcal{C} = \left\{ c_{\mathbf{p},f} : c_{\mathbf{p},f}(\mathbf{r}) = \sum_{i\in\mathcal{N}} \frac{r_i}{p_i} - (1-f),\ \mathbf{p} \in \mathcal{P},\ f \in \mathcal{T}_{fr} \right\}$$

where $\mathcal{T}_{fr}$ is a finite set of real numbers in [0, 1) where each element in the set corresponds to a fraction of a time slot that is allocated to other traffic.

3) Time-Varying Utility Functions: Additionally, our framework also allows the introduction of time-varying utility functions as illustrated by the following example of a wireless network supporting video users. Here, we view utility functions as a mapping from allocated resource (e.g., rate) to reward (e.g., video quality). For video users, we consider perceived video quality of a user in a time slot as the reward for that user in that slot. However, for video users, the dependence of perceived video quality[3] on the compression rate is time varying. This is typically due to the possibly changing nature of the content, e.g., from an action to a slower scene. Hence, the utility function that maps the reward (i.e., perceived video quality) derived from the allocated resource (i.e., the rate) is time varying. This setting can be handled as follows. Let $q_{t,i}(\cdot)$ denote the strictly increasing concave function that, in time slot t, maps the rate allocated to user i to user perceived video quality. For each user i, let $\mathcal{Q}_i$ be a finite set of such functions, then a scenario with time varying peak rates and utilities can be modeled by set of convex constraints

$$\mathcal{C} = \left\{ c_{\mathbf{p},\mathbf{q}} : c_{\mathbf{p},\mathbf{q}}(\mathbf{r}) = \sum_{i\in\mathcal{N}} \frac{q_i^{-1}(r_i)}{p_i} - 1, \mathbf{p} \in \mathcal{P}, q_i \in \mathcal{Q}_i\ \forall\ i \in \mathcal{N} \right\}.$$

[3]In a short duration time slot roughly a second long which corresponds to a collection of 20–30 frames.

III. Optimal Variance-Sensitive Offline Policy

In this section, we study OPT(T), the offline formulation for optimal reward allocation introduced in Section I. In the offline setting, we assume that $(c)_{1:T}$, the realization of the constraints process $(C)_{1:T}$, is known. We denote the objective function of OPT(T) by ϕ_T, i.e.,

$$\phi_T(\mathbf{r}) := \sum_{i\in\mathcal{N}} U_i^E\left(e_i^T(r_i)\right) \tag{5}$$

where $e_i^T(\cdot)$ is as in Definition 1. Hence the optimization problem $\mathrm{OPT}(T)$ can be rewritten as

$$\mathbf{OPT}(T): \max_{(\mathbf{r})_{1:T}} \quad \phi_T(\mathbf{r}) \tag{6}$$

$$\text{subject to} \quad c_t\left(\mathbf{r}(t)\right) \leq 0\ \ \forall\, t \in \{1,\ldots,T\} \tag{7}$$

$$r_i(t) \geq 0\ \ \forall\, t \in \{1,\ldots,T\},\ \forall\, i \in \mathcal{N}. \tag{8}$$

The next result asserts that $\mathrm{OPT}(T)$ is a convex optimization problem satisfying Slater's condition [6, Section 5.2.3] and that it has a unique solution.

Lemma 1: $\mathrm{OPT}(T)$ is a convex optimization problem satisfying Slater's condition with a unique solution.

Proof: By Assumptions U.E and U.V, the convexity of the objective of $\mathrm{OPT}(T)$ is easy to establish once we prove the convexity of the function $U_i^V(\mathrm{Var}^T(\cdot))$ for each $i \in \mathcal{N}$. Using (3) and the definition of $\mathrm{Var}^T(\cdot)$, we can show that $U_i^V(\mathrm{Var}^T(\cdot))$ is convex for each $i \in \mathcal{N}$. The details are given next. Using convexity of Euclidean norm (see [6]), we can show that for any two quality vectors $(\mathbf{r^1})_{1:T}$ and $(\mathbf{r^2})_{1:T}$, any $i \in \mathcal{N}$, $\alpha \in (0,1)$ and $\bar{\alpha} = 1 - \alpha$, we have that

$$\mathrm{Var}^T\left(\alpha r_i^1 + \bar{\alpha} r_i^2\right) = \left(\alpha\sqrt{\mathrm{Var}^T\left(r_i^1\right)} + \bar{\alpha}\sqrt{\mathrm{Var}^T\left(r_i^2\right)}\right)^2. \tag{9}$$

Using this, (3) and the monotonicity of U_i^V, we have

$$U_i^V\left(\mathrm{Var}^T\left(\alpha r_i^1 + \bar{\alpha} r_i^2\right)\right) \leq \alpha U_i^V\left(\mathrm{Var}^T\left(r_i^1\right)\right) + \bar{\alpha} U_i^V\left(\mathrm{Var}^T\left(r_i^2\right)\right). \tag{10}$$

So, $U_i^V(\mathrm{Var}^T(\cdot))$ is a convex function. Thus, by the concavity of $U_i^E(\cdot)$ and $-U_i^V(\mathrm{Var}^T(\cdot))$, we can conclude that $\mathrm{OPT}(T)$

is a convex optimization problem. Also, from (9) and (3) (since we have strict inequality), we can conclude that we have equality in (10) only if

$$\mathrm{Var}^T\left(r_i^1\right) = \mathrm{Var}^T\left(r_i^2\right) \tag{11}$$

or equivalently

$$r_i^1(t) = r_i^2(t) + m^T\left(r_i^1\right) - m^T\left(r_i^2\right) \ \ \forall t \in \{1,\ldots,T\}. \tag{12}$$

Further, Slater's condition is satisfied and it follows from (2) in Assumption C.3.

Now, for any $i \in \mathcal{N}$, U_i^E and $-U_i^V(\mathrm{Var}^T(\cdot))$ are not necessarily strictly concave. But, we can still show that $\mathrm{OPT}(T)$ has a unique solution. Let $(\mathbf{r^1})_{1:T}$ and $(\mathbf{r^2})_{1:T}$ be two optimal solutions to $\mathrm{OPT}(T)$. Then, from the concavity of the objective, $(\alpha(r_i^1)_{1:T} + \bar{\alpha}(r_i^2)_{1:T})$ is also an optimal solution for any $\alpha \in (0,1)$ and $\bar{\alpha} = 1-\alpha$. Due to convexity of $U_i^E(\cdot)$ and $U_i^V(\mathrm{Var}^T s(\cdot))$, this is only possible if for each $i \in \mathcal{N}$ and $1 \le t \le T$

$$U_i^V\left(\mathrm{Var}^T\left(\alpha r_i^1 + \bar{\alpha} r_i^2\right)\right) = \alpha U_i^V\left(\mathrm{Var}^T\left(r_i^1\right)\right) + \bar{\alpha} U_i^V\left(\mathrm{Var}^T\left(r_i^2\right)\right).$$

Hence (12) and (11) hold. Due to optimality of $(\mathbf{r^1})_{1:T}$ and $(\mathbf{r^2})_{1:T}$, we have that

$$\begin{aligned}
&\sum_{i\in\mathcal{N}} U_i^E\left(\frac{1}{T}\sum_{t=1}^T r_i^2(t) - U_i^V\left(\mathrm{Var}^T\left(r_i^2\right)\right)\right)\\
&\quad= \sum_{i\in\mathcal{N}} U_i^E\left(\frac{1}{T}\sum_{t=1}^T r_i^1(t) - U_i^V\left(\mathrm{Var}^T\left(r_i^2\right)\right)\right)\\
&\quad= \sum_{i\in\mathcal{N}} U_i^E\left(\frac{1}{T}\sum_{t=1}^T r_i^2(t) + m^T\left(r_i^1\right) - m^T\left(r_i^2\right)\right.\\
&\qquad\qquad\left. - U_i^V\left(\mathrm{Var}^T\left(r_i^2\right)\right)\right)
\end{aligned}$$

where the first equality follows from (11) and the second one follows from (12). Since U_i^E is a strictly increasing function for each $i \in \mathcal{N}$, the above equation implies that $m^T(r_i^1) = m^T(r_i^2)$ and thus (using (12)) $\mathbf{r^1}(t) = \mathbf{r^2}(t)$ for each t such that $1 \le t \le T$. From the above discussion, we can conclude that $\mathrm{OPT}(T)$ has a unique solution. ■

We let $(\mathbf{r}^T)_{1:T}$ denote the optimal solution to $\mathrm{OPT}(T)$. Since $\mathrm{OPT}(T)$ is a convex optimization problem satisfying Slater's condition (Lemma 1), the Karush-Kuhn-Tucker (KKT) conditions (see [6, Section 5.5.3]) given next hold.

KKT-OPT(T):

There exist nonnegative constants $(\mu^T)_{1:T}$ and $(\boldsymbol{\gamma}^T)_{1:T}$ such that for all $i \in \mathcal{N}$ and $t \in \{1,\ldots,T\}$, we have

$$\left(U_i^E\right)'\left(e_i^T\left(r_i^T\right)\right)\left(\frac{1}{T} - \frac{2\left(U_i^V\right)'\left(\mathrm{Var}^T\left(r_i^T\right)\right)}{T}\left(r_i^T(t) - m^T\left(r_i^T\right)\right)\right)$$

$$- \frac{\mu^T(t)}{T} c'_{t,i}\left(\mathbf{r}^T(t)\right) + \frac{\gamma_i^T(t)}{T} = 0 \tag{13}$$

$$\mu^T(t) c_t\left(\mathbf{r}^T(t)\right) = 0 \tag{14}$$

$$\gamma_i^T(t) r_i^T(t) = 0 \tag{15}$$

Here $c'_{t,i}$ denotes $\partial c_t/\partial r_i$, and we have used the fact that for any $t \in \{1,\ldots,T\}$

$$\frac{\partial}{\partial r(t)}\left(T\mathrm{Var}^T(r)\right) = 2\left(r(t) - m^T(r)\right).$$

From (13), we see that the optimal reward allocation $\mathbf{r}^T(t)$ on time slot t depends on the entire allocation $(\mathbf{r^T})_{1:T}$ through the following three quantities: (i) the time average rewards $\mathbf{m}^T$; (ii) $\left(\left(U_i^E\right)'\right)_{i\in\mathcal{N}}$ evaluated at the quality of experience of the respective users; and (iii) $\left(\left(U_i^V\right)'\right)_{i\in\mathcal{N}}$ evaluated at the variance seen by the respective users. So, if the time averages associated with the *optimal solution* were somehow known, the optimal allocation for each time slot t could be determined by solving an optimization problem (derived from the KKT conditions) that only requires these time averages, and knowledge of c_t (associated with current time slot) rather than $(c)_{1:T}$. We exploit this key idea in formulating our online algorithm in the next section.

IV. Adaptive Variance-Aware Reward Allocation

In this section, we present Adaptive Variance-aware Reward allocation (AVR) algorithm and establish its asymptotic optimality.

We let

$$\mathcal{H} := [0, r_{\max}]^N \times [0, v_{\max}]^N \tag{16}$$

where $\times$ denotes Cartesian product for sets. Let $(\mathbf{m}, \boldsymbol{v}) \in \mathcal{H}$ and $e_i = m_i - U_i^V(v_i)$ for each $i \in \mathcal{N}$, and consider the optimization problem $\mathrm{OPTAVR}((\mathbf{m},\boldsymbol{v}),c)$ given below

---

OPTAVR$((\mathbf{m},\boldsymbol{v}),c)$:

$$\max_{\mathbf{r}} \sum_{i\in\mathcal{N}} \left(U_i^E\right)'(e_i)\left(r_i - \left(U_i^V\right)'(v_i)(r_i - m_i)^2\right)$$

$$\text{subject to} \quad c(\mathbf{r}) \le 0 \tag{17}$$

$$r_i \ge 0 \quad \forall\, i \in \mathcal{N}. \tag{18}$$

The reward allocations for AVR are obtained by solving $\mathrm{OPTAVR}((\mathbf{m},\boldsymbol{v}),c)$, where $\mathbf{m}$, $\boldsymbol{v}$, and $\mathbf{e}$ correspond to current estimates of the mean, variance and QoE, respectively. We let $\mathbf{r}^*((\mathbf{m},\boldsymbol{v}),c)$ denote the optimal solution to $\mathrm{OPTAVR}((\mathbf{m},\boldsymbol{v}),c)$.

Next, we describe our algorithm in detail.

Algorithm 1. Adaptive Variance-aware Reward allocation (AVR)

AVR.0: Initialization: let $(\mathbf{m}(1), \boldsymbol{v}(1)) \in \mathcal{H}$.

In each time slot $t \in \mathbb{N}$, carry out the following steps:

AVR.1: The reward allocation in time slot t is the optimal solution to $\text{OPTAVR}((\mathbf{m}(t), \boldsymbol{v}(t)), c_t)$, i.e., $\mathbf{r}^*((\mathbf{m}(t), \boldsymbol{v}(t)), c_t)$, and will be denoted by $\mathbf{r}^*(t)$ (when the dependence on the variables is clear from context).

AVR.2: In time slot t, update m_i as follows: for all $i \in \mathcal{N}$

$$m_i(t+1) = \left[m_i(t) + \frac{1}{t}\left(r_i^*(t) - m_i(t)\right) \right]_0^{r_{\max}} \tag{19}$$

and update v_i as follows: for all $i \in \mathcal{N}$

$$v_i(t+1) = \left[v_i(t) + \frac{\left(r_i^*(t) - m_i(t)\right)^2 - v_i(t)}{t} \right]_0^{v_{\max}}. \tag{20}$$

Here, $[x]_a^b = \min(\max(x, a), b)$.

Thus, AVR greedily allocates rewards in slot t based on the objective of $\text{OPTAVR}((\mathbf{m}(t), \boldsymbol{v}(t)), c_t)$. Thus, the computational requirements per slot involve solving a convex program in N variables (that has a simple quadratic function as its objective function), and updating at most $2N$ variables. We see that the update equations (19), (20) roughly ensure that the parameters $\mathbf{m}(t)$ and $\boldsymbol{v}(t)$ keep track of mean reward and variance in reward allocations under AVR. The updates in AVR fall in the class of decreasing step size stochastic approximation algorithms (see [17] for reference) due to the use of $1/t$ in (19), (20). We could replace $1/t$ with a small positive constant ϵ and obtain a constant step size stochastic approximation algorithm which is usually better suited for non-stationary settings (also see [27, Section 4.4] for other useful choices). Note that we do not have to keep track of variance estimates for users i with linear U_i^V since OPTAVR is insensitive to their values (i.e., $(U_i^V)'(.)$ is a constant), and thus the evolutions of $\mathbf{m}(t)$ and $(v_i(t))_{i \in \mathcal{N}_n}$ do not depend on them. We let $\boldsymbol{\theta}(t) = (\mathbf{m}(t), \boldsymbol{v}(t))$ for each t. The truncation $[.]_a^b$ in the update (19), (20) ensure that $\boldsymbol{\theta}(t)$ stays in the set $\mathcal{H}$.

For any $((\mathbf{m}, \boldsymbol{v}), c) \in \mathcal{H} \times \mathcal{C}$, we have $(U_i^E)'(m_i - U_i^V(v_i))(U_i^V)'(v_i) > 0$ for each $i \in \mathcal{N}$ (see Assumptions U.E and U.V). Hence, $\text{OPTAVR}((\mathbf{m}, \boldsymbol{v}), c)$ is a convex optimization problem with a unique solution. Further, using (2) in Assumption C.3, we can show that it satisfies Slater's condition. Hence, the optimal solution $\mathbf{r}^*$ for $\text{OPTAVR}((\mathbf{m}, \boldsymbol{v}), c)$ satisfies KKT conditions given below.

KKT-OPTAVR$((\mathbf{m}, \boldsymbol{v}), c)$:

There exist nonnegative constants μ^* and $(\gamma_i^*)_{i \in \mathcal{N}}$ such that for all $i \in \mathcal{N}$

$$\left(U_i^E\right)'\left(m_i - U_i^V(v_i)\right)\left(1 - 2\left(U_i^V\right)'(v_i)\left(r_i^* - m_i\right)\right) + \gamma_i^* - \mu^* c_i'(\mathbf{r}^*) = 0 \tag{21}$$

$$\mu^* c(\mathbf{r}^*) = 0 \tag{22}$$

$$\gamma_i^* r_i^* = 0. \tag{23}$$

In the next lemma, we establish continuity properties of $\mathbf{r}^*((\mathbf{m}, \boldsymbol{v}), c)$ when viewed as a function of $(\mathbf{m}, \boldsymbol{v})$. In particular, the Lipschitz assumption on the derivatives of $(U_i^V)_{i \in \mathcal{N}}$ and $(U_i^E)_{i \in \mathcal{N}}$ help us conclude that the optimizer of $\text{OPTAVR}(\boldsymbol{\theta}, c)$ is Lipschitz continuous in $\boldsymbol{\theta}$. A proof is given in Appendix A.

Lemma 2: For any $c \in \mathcal{C}$, and $\boldsymbol{\theta} = (\mathbf{m}, \boldsymbol{v}) \in \mathcal{H}$

(a) $\mathbf{r}^*(\boldsymbol{\theta}, c)$ is a Lipschitz continuous function of $\boldsymbol{\theta}$.

(b) $E[\mathbf{r}^*(\boldsymbol{\theta}, C^\pi)]$ is a Lipschitz continuous function of $\boldsymbol{\theta}$.

The next theorem states our key convergence result for the mean, variance and QoE of the reward allocations under AVR. This result is proven in Section V. For brevity, we let $\mathbf{r}^*(t)$ denote $\mathbf{r}^*((\mathbf{m}(t), \boldsymbol{v}(t)), c_t)$.

Theorem 1: The evolution of the users' estimated parameters $\mathbf{m}(t)$ and $\mathbf{v}(t)$, and the sequence of reward allocations $(r_i^*)_{1:T}$ to each user i under AVR satisfy the following property: for almost all sample paths, and for each $i \in \mathcal{N}$

$$(a) \quad \lim_{T\to\infty} m^T\left(r_i^*\right) = \lim_{t\to\infty} m_i(t)$$

$$(b) \quad \lim_{T\to\infty} \text{Var}^T\left(r_i^*\right) = \lim_{t\to\infty} v_i(t)$$

$$(c) \quad \lim_{T\to\infty} e_i^T\left(r_i^*\right) = \lim_{t\to\infty} \left(m_i(t) - U_i^V\left(v_i(t)\right)\right).$$

The next result establishes the asymptotic optimality of AVR, i.e., if we consider long periods of time T, the difference in performance (i.e., ϕ_T defined in (5)) of the online algorithm AVR and the optimal offline policy $\text{OPT}(T)$ becomes negligible. Thus, the sum utility of the QoEs (which depends on long term time averages) is optimized.

Theorem 2: The sequence of reward allocations $(\mathbf{r}^*)_{1:T}$ under AVR is feasible, i.e., it satisfies (7) and (8), and for almost all sample paths they are asymptotically optimal, i.e.,

$$\lim_{T\to\infty} \left(\phi_T(\mathbf{r}^*) - \phi_T(\mathbf{r}^T)\right) = 0.$$

Proof: Since the allocation $(\mathbf{r}^*)_{1:T}$ associated with AVR satisfies (17) and (18) at each time slot, it also satisfies (7) and (8).

To show asymptotic optimality, consider any realization of $(c)_{1:T}$. Let $(\mu^*)_{1:T}$ and $(\boldsymbol{\gamma}^*)_{1:T}$ be the sequences of nonnegative real numbers satisfying (21)–(23) for this realization. From the nonnegativity of these numbers, and feasibility of $(\mathbf{r}^T)_{1:T}$, we have

$$\phi_T(\mathbf{r}^T) \le \psi_T(\mathbf{r}^T) \tag{24}$$

where

$$\psi_T(\mathbf{r}^T) = \sum_{i\in\mathcal{N}} U_i^E\left(e_i^T\left(r_i^T\right)\right) - \sum_{t=1}^{T} \frac{\mu^*(t)}{T} c_t\left(\mathbf{r}^T(t)\right) + \sum_{t=1}^{T}\sum_{i\in\mathcal{N}} \frac{\gamma_i^*(t)}{T} r_i^T(t).$$

Indeed, the function ψ_T is the Lagrangian associated with $\mathrm{OPT}(T)$ but evaluated at the optimal Lagrange multipliers associated with the optimization problems (OPTAVR) involved in AVR, and hence the inequality. Since ψ_T is a differentiable concave function, we have (see [6])

$$\psi_T(\mathbf{r}^T) \leq \psi_T(\mathbf{r}^*) + \left\langle \nabla\psi_T(\mathbf{r}^*), \left((\mathbf{r}^T)_{1:T} - (\mathbf{r}^*)_{1:T}\right)\right\rangle$$

where $\langle\cdot,\cdot\rangle$ denotes the dot product. Hence, we have

$$\begin{aligned}\psi_T(\mathbf{r}^T) \leq & \sum_{i\in\mathcal{N}} U_i^E\left(e_i^T\left(r_i^*\right)\right) - \sum_{t=1}^{T}\frac{\mu^*(t)}{T}c_t\left(\mathbf{r}^*(t)\right)\\ &+\sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\frac{\gamma_i^*(t)}{T}r_i^*(t) + \sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\left(r_i^T(t) - r_i^*(t)\right)\\ &\left(-\frac{\mu^*(t)}{T}c'_{t,i}\left(\mathbf{r}^*(t)\right) + \frac{\gamma_i^*(t)}{T} + \left(U_i^E\right)'\left(e_i^T\left(r_i^*\right)\right)\right.\\ &\left.\left(\frac{1}{T} - \frac{2\left(U_i^V\right)'\left(\mathrm{Var}^T\left(r_i^*\right)\right)}{T}\left(r_i^*(t) - m^T\left(r_i^*\right)\right)\right)\right).\end{aligned}$$

Using (24), and the fact that $(\mu^*)_{1:T}$ and $(\boldsymbol{\gamma}^*)_{1:T}$ satisfy (21)–(23), we have

$$\begin{aligned}\phi_T(\mathbf{r}^T) \leq & \sum_{i\in\mathcal{N}} U_i^E\left(e_i^T\left(r_i^*\right)\right) + \sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\frac{r_i^T(t) - r_i^*(t)}{T}\\ &\left(\left(U_i^E\right)'\left(e_i^T\left(r_i^*\right)\right)\right.\\ &\left.\left(1 - 2\left(U_i^V\right)'\left(\mathrm{Var}^T\left(r_i^*\right)\right)\left(r_i^*(t) - m^T\left(r_i^*\right)\right)\right)\right.\\ &-\left(U_i^E\right)'\left(e_i(t-1)\right)\\ &\left.\left(1 - 2\left(U_i^V\right)'\left(v_i(t-1)\right)\left(r_i^*(t) - m_i(t-1)\right)\right)\right). \quad (25)\end{aligned}$$

From Theorem 1 (a)–(c), and the continuity and boundedness of the functions involved, we can conclude that the expression appearing in the last four lines of the above inequality can be made as small as desired by choosing large enough T and then choosing a large enough t. Also, $|r_i^T(t) - r_i^*(t)| \leq r_{\max}$ for each $i \in \mathcal{N}$. Hence, taking limits in (25)

$$\lim_{T\to\infty}\left(\phi_T(\mathbf{r}^*) - \phi_T(\mathbf{r}^T)\right) \geq 0 \quad (26)$$

holds for almost all sample paths. From the optimality of $(\mathbf{r}^T)_{1:T}$

$$\phi_T(\mathbf{r}^T) \geq \phi_T(\mathbf{r}^*). \quad (27)$$

The result follows from the inequalities (26) and (27). ■

Remark: The asymptotic optimality of AVR (stated in Theorem 2) can also be established when $(C_t)_{t\in\mathbb{N}}$ is an asymptotically mean stationary process (see [11] for reference) of functions selected from a finite set $\mathcal{C}$. This is a weaker assumption than Assumption C.1, and our proofs of Lemma 7 and Theorem 1 used Assumption C.1. Note that the proof of Lemma 7 (given in Appendix G) holds as long as the strong law of large numbers holds for $(g(\mathbf{m},\boldsymbol{v},C_t))_t$ for any $(\mathbf{m},\boldsymbol{v})$, and the proof of Theorem 1 can be extended as long as the second term in (42) converges to $\mathbb{E}[r_i^*(\boldsymbol{\theta}^\pi, C^\pi)]$. Note that both these modifications to the proofs hold since Birkhoff's Ergodic Theorem (BET) holds for $(C_t)_{t\in\mathbb{N}}$ (see paragraph below Remark 7 in Section 3 of [11]).

V. Convergence Analysis

This section is devoted to the proof of the previously stated Theorem 1 capturing the convergence of reward allocations under AVR. Our approach relies on viewing (19), (20) in AVR as a stochastic approximation update equation (see, e.g., [17] for reference), and relating the convergence of reward allocations under the discrete time algorithm AVR to that of an auxiliary (continuous time) ODE (given in (37)) which evolves according to time averaged dynamics of AVR. In fact, we will show that the ODE converges to a point determined by the optimal solution to an auxiliary optimization problem OPTSTAT closely related to $\mathrm{OPT}(T)$ which is discussed in the next subsection. In Section V-B, we study the convergence of the auxiliary ODE and in Section V-C, we establish convergence of $(\boldsymbol{\theta}(t))_{t\in\mathbb{N}}$ generated by AVR to complete the proof of Theorem 1.

A. Stationary Version of OPT: OPTSTAT

The formulation OPT(T) involves time averages of various quantities associated with users' rewards. By contrast, the formulation of OPTSTAT is based on expected values of the corresponding quantities under the stationary distribution of $(C_t)_{t\in\mathbb{N}}$.

Recall that (under Assumption C.1) $(C_t)_{t\in\mathbb{N}}$ is a stationary ergodic process with marginal distribution $(\pi(c))_{c\in\mathcal{C}}$, i.e., for $c \in \mathcal{C}$, $\pi(c)$ is the probability of the event $C_t = c$. Since $\mathcal{C}$ is finite, we assume that $\pi(c) > 0$ for each $c \in \mathcal{C}$ without any loss of generality.

Definition 2: A reward allocation policy is said to be *stationary* if the associated reward allocation in any time slot t depends only on current constraint c_t.

Thus, we can represent any stationary reward allocation policy as a $|\mathcal{C}|$ length vector (of vectors) $(\boldsymbol{\rho}_c)_{c\in\mathcal{C}}$ where $\boldsymbol{\rho}_c = (\rho_{c,i})_{i\in\mathcal{N}} \in \mathbb{R}_+^N$ denotes the allocation of rewards to users under constraint $c \in \mathcal{C}$.

Definition 3: We say that a stationary reward allocation policy $(\boldsymbol{\rho}_c)_{c\in\mathcal{C}}$ is *feasible* if for each $c \in \mathcal{C}$, we have that $c(\boldsymbol{\rho}_c) \leq 0$ and for each $i \in \mathcal{N}$, we have $\rho_{c,i} \geq 0$. Also, let $\mathcal{R}_\mathcal{C} \subset \mathbb{R}^{N|\mathcal{C}|}$ denote the set of feasible stationary reward allocation policies, i.e.,

$$\mathcal{R}_\mathcal{C} := \Pi_{c\in\mathcal{C}}\left\{\boldsymbol{\rho}_c \in \mathbb{R}^N : c(\boldsymbol{\rho}_c) \leq 0,\ \rho_{c,i} \geq 0 \quad \forall\, i \in \mathcal{N}\right\}. \quad (28)$$

Now, let

$$\phi_\pi\left((\boldsymbol{\rho}_c)_{c\in\mathcal{C}}\right) = \sum_{i\in\mathcal{N}} U_i^E\left(E[\rho_{C^\pi,i}] - U_i^V\left(\mathrm{Var}(\rho_{C^\pi,i})\right)\right)$$

where $\rho_{C^\pi,i}$ is a random variable taking value $\rho_{c,i}$ with probability $\pi(c)$ for each $c \in \mathcal{C}$, i.e., a random variable whose

distribution is that of user i's reward allocation under stationary reward allocation policy $(\boldsymbol{\rho}_c)_{c\in\mathcal{C}}$. Hence

$$E[\rho_{C^\pi,i}] = \sum_{c\in\mathcal{C}} \pi(c)\rho_{c,i}$$

$$\mathrm{Var}(\rho_{C^\pi,i}) = \sum_{c\in\mathcal{C}} \pi(c)\left(\rho_{c,i} - E[\rho_{C^\pi,i}]\right)^2.$$

We define the 'stationary' optimization problem OPTSTAT as follows:

OPTSTAT:

$$\max_{(\boldsymbol{\rho}_c)_{c\in\mathcal{C}}\in\mathcal{R}_\mathcal{C}} \quad \phi_\pi\left((\boldsymbol{\rho}_c)_{c\in\mathcal{C}}\right).$$

The next lemma gives a few useful properties of OPTSTAT.

Lemma 3: OPTSTAT is a convex optimization problem satisfying Slater's condition and has a unique solution.

Proof: The proof is similar to that of Lemma 1, and is easy to establish once the convexity of the function $\mathrm{Var}(\cdot)$ is shown. ■

Using Lemma 3, we can conclude that the KKT conditions given below are necessary and sufficient for optimality of OPTSTAT. Let $(\boldsymbol{\rho}_c^\pi)_{c\in\mathcal{C}}$ denote the optimal solution.

KKT-OPTSTAT:

There exist constants $(\mu^\pi(c))_{c\in\mathcal{C}}$ and $(\boldsymbol{\gamma}^\pi(c))_{c\in\mathcal{C}}$ such that

$$\pi(c)\left(U_i^E\right)'\left(E\left[\rho^\pi_{C^\pi,i}\right] - U_i^V\left(\mathrm{Var}\left(\rho^\pi_{C^\pi,i}\right)\right)\right)$$
$$\left(1 - 2\left(U_i^V\right)'\left(\mathrm{Var}\left(\rho^\pi_{C^\pi,i}\right)\right)\left(\rho^\pi_{c,i} - E\left[\rho^\pi_{C^\pi,i}\right]\right)\right)$$
$$- \mu^\pi(c)c_i'\left(\boldsymbol{\rho}_c^\pi\right) + \gamma_i^\pi(c) = 0 \quad (29)$$

$$\mu^\pi(c)c\left(\boldsymbol{\rho}_c^\pi\right) = 0 \quad (30)$$

$$\gamma_i^\pi(c)\rho^\pi_{c,i} = 0 \quad (31)$$

where c_i' denotes the ith component of the gradient ∇c of the constraint function $c \in \mathcal{C}$.

In developing the above KKT conditions, we used the fact that for any $c \in \mathcal{C}$ and $i \in \mathcal{N}$, $\partial\mathrm{Var}(\rho^\pi_{C^\pi,i})/\partial\rho_{c,i} = 2\pi(c)(\rho^\pi_{c,i} - E[\rho^\pi_{C^\pi,i}])$.

Next, we find relationships between the optimal solution $(\boldsymbol{\rho}_c^\pi)_{c\in\mathcal{C}}$ of OPTSTAT and OPTAVR. To that end, let $\boldsymbol{\theta}^\pi := (\mathbf{m}^\pi, \mathbf{v}^\pi)$ where for each $i \in \mathcal{N}$, we define

$$m_i^\pi := E\left[\rho^\pi_{C^\pi,i}\right] \quad (32)$$

$$v_i^\pi := \mathrm{Var}^\pi\left(\rho^\pi_{C^\pi,i}\right) \quad (33)$$

$$e_i^\pi := m_i^\pi - U_i^V\left(v_i^\pi\right). \quad (34)$$

Definition 4: Let $\mathcal{H}^*$ be the set of fixed points defined by

$$\mathcal{H}^* = \{(\mathbf{m},\boldsymbol{v}) \in \mathcal{H} : (\mathbf{m},\boldsymbol{v}) \text{ satisfies (35)–(36)}\}$$

where

$$E\left[r_i^*\left((\mathbf{m},\boldsymbol{v}), C^\pi\right)\right] = m_i \quad \forall\, i \in \mathcal{N} \quad (35)$$

$$\mathrm{Var}\left(r_i^*\left((\mathbf{m},\boldsymbol{v}), C^\pi\right)\right) = v_i \quad \forall\, i \in \mathcal{N}. \quad (36)$$

Recall that $\mathbf{r}^*((\mathbf{m},\boldsymbol{v}),c)$ denotes the optimal solution to OPTAVR$((\mathbf{m},\boldsymbol{v}),c)$ and $\mathcal{H}$ is defined in (16). Thus, $\mathcal{H}^*$ is the set of parameter values $\boldsymbol{\theta} = (\mathbf{m},\boldsymbol{v})$ that can be viewed as fixed points for 'stationary modification' of AVR obtained by replacing $r_i^*(t)$ and $(r_i^*(t) - m_i(t))^2$ in (19) and (20) with their expected values. Theorem 3 below shows that in fact there is but one such fixed point $\boldsymbol{\theta}^\pi$. A proof is given in Appendix B.

Theorem 3: $\boldsymbol{\theta}^\pi$ satisfies the following:

(a) $\mathbf{r}^*(\boldsymbol{\theta}^\pi, c) = \boldsymbol{\rho}_c^\pi$ for each $c \in \mathcal{C}$, and

(b) $\mathcal{H}^* = \{\boldsymbol{\theta}^\pi\}$.

Using these results we will study a differential equation that mimics the evolution of the parameters under AVR and show that it converges to $\boldsymbol{\theta}^\pi$.

B. Convergence of Auxiliary ODE Associated With AVR

In this subsection, we study and establish convergence of an auxiliary ODE which evolves according to the average dynamics of AVR. We establish the relationship between the ODE and AVR in the next subsection. This will subsequently be used in establishing convergence properties of AVR.

Consider the following differential equation:

$$\frac{d\boldsymbol{\theta}^A(\tau)}{d\tau} = \bar{\mathbf{g}}\left(\boldsymbol{\theta}^A(\tau)\right) + \mathbf{z}\left(\boldsymbol{\theta}^A(\tau)\right) \quad (37)$$

for $\tau \geq 0$ with $\boldsymbol{\theta}^A(0) \in \mathcal{H}$ where $\bar{\mathbf{g}}(\boldsymbol{\theta})$ is a function taking values in $\mathbb{R}^{2N}$ defined as follows: for $\boldsymbol{\theta} = (\mathbf{m},\boldsymbol{v}) \in \mathcal{H}$, let

$$\left(\bar{\mathbf{g}}(\boldsymbol{\theta})\right)_i := E\left[r_i^*(\boldsymbol{\theta}, C^\pi)\right] - m_i \quad (38)$$

$$\left(\bar{\mathbf{g}}(\boldsymbol{\theta})\right)_{N+i} := E\left[\left(r_i^*(\boldsymbol{\theta}, C^\pi) - m_i\right)^2\right] - v_i. \quad (39)$$

In (37), $z(\boldsymbol{\theta}) \in -C_\mathcal{H}(\boldsymbol{\theta})$ is a projection term corresponding to the smallest vector that ensures that the solution remains in $\mathcal{H}$ (see [17, Section 4.3]). The set $C_\mathcal{H}(\boldsymbol{\theta})$ contains only the zero element when $\boldsymbol{\theta}$ is in the interior of $\mathcal{H}$, and for $\boldsymbol{\theta}$ on the boundary of the set $\mathcal{H}$, $C_\mathcal{H}(\boldsymbol{\theta})$ is the convex cone generated by the outer normals at $\boldsymbol{\theta}$ of the faces of $\mathcal{H}$ on which $\boldsymbol{\theta}$ lies. The motivation for studying the above differential equation should be partly clear by comparing the right hand side of (37) (see (38), (39)) with AVR's update (19), (20), and we can associate the term $z(\boldsymbol{\theta})$ with the constrained nature of AVR's update equations. The following result shows that $z(\boldsymbol{\theta})$ appearing in (37) is innocuous in the sense that we can ignore it when we study the differential equation. The proof, given in Appendix C, shows the redundancy of the term $z(\boldsymbol{\theta})$ by arguing that the differential equation itself ensures that $\boldsymbol{\theta}^A(\tau)$ stays within $\mathcal{H}$.

Lemma 4: For any $\boldsymbol{\theta} \in \mathcal{H}$, $z_j(\boldsymbol{\theta}) = 0$ for all $1 \leq j \leq 2N$.

Note that (37) has a unique solution for a given initialization due to Lipschitz continuity results in Lemma 2.

We define the set $\widetilde{\mathcal{H}} \subset \mathcal{H}$ as follows:

$$\widetilde{\mathcal{H}} := \big\{(\mathbf{m},\boldsymbol{v}) \in \mathcal{H} : \text{ there exists } (\boldsymbol{\rho}_c)_{c\in\mathcal{C}} \in \mathcal{R}_\mathcal{C} \text{ such that}$$
$$E[\rho_{C^\pi,i}] = m_i,\ \mathrm{Var}(\rho_{C^\pi,i}) \leq v_i \leq r_{\max}^2 \quad \forall\, i \in \mathcal{N}\big\}$$

where $\mathcal{R}_{\mathcal{C}}$ is the set of feasible stationary reward allocation policies defined in (28). We can view $\widetilde{\mathcal{H}}$ as the set of all 'achievable' mean variance pairs, i.e., for any $(\mathbf{m}, \boldsymbol{v}) \in \mathcal{H}$ there is some stationary allocation policy with associated mean vector equal to $\mathbf{m}$ and associated variance vector componentwise less than or equal to $\boldsymbol{v}$. Here, the restriction $v_i \leq r_{\max}^2$ for each i ensures that $\widetilde{\mathcal{H}}$ is bounded. Further, for any $\boldsymbol{\theta} = (\mathbf{m}, \boldsymbol{v}) \in \widetilde{\mathcal{H}}$, let

$$\widetilde{\mathcal{R}}(\boldsymbol{\theta}) := \left\{ (\boldsymbol{\rho}_c)_{c\in\mathcal{C}} \in \mathcal{R}_{\mathcal{C}} : E[\rho_{C^\pi,i}] = m_i, \right. \\ \left. \mathrm{Var}(\rho_{C^\pi,i}) \leq v_i \quad \forall\, i \in \mathcal{N} \right\}.$$

We can view $\widetilde{\mathcal{R}}(\boldsymbol{\theta})$ as the set of all feasible stationary reward allocation policies corresponding to an achievable $\boldsymbol{\theta} \in \widetilde{\mathcal{H}}$.

The following result characterizes several useful properties of the sets introduced above; a proof is given in Appendix D.

Lemma 5: (a) For any $\boldsymbol{\theta} = (\mathbf{m}, \boldsymbol{v}) \in \widetilde{\mathcal{H}}$, $\widetilde{\mathcal{R}}(\boldsymbol{\theta})$ is a non-empty compact subset of $\mathbb{R}^{N|\mathcal{C}|}$.

(b) $\widetilde{\mathcal{H}}$ is a bounded, closed and convex set.

The next result gives a set of sufficient conditions to establish asymptotic stability of a point with respect to an ordinary differential equation. This result is a generalization of Theorem 4 in [30]. A proof of the result is given in Appendix E.

Lemma 6: Consider a differential equation

$$\dot{x} = f(x), \qquad x \in \mathbb{R}^d \tag{40}$$

where f is locally Lipschitz and all trajectories exist for $t \in [0, \infty)$. Suppose that some compact set $K \subset \mathbb{R}^d$ is asymptotically stable with respect to (40) and also suppose that there exists a continuously differentiable function $L : \mathbb{R}^d \to \mathbb{R}$ and some $x_0 \in K$ such that

$$\nabla L(x) \cdot f(x) < 0 \qquad \forall\, x \in K,\ x \neq x_0. \tag{41}$$

Then x_0 is an asymptotically stable equilibrium for (40) in $\mathbb{R}^d$.

We are now in a position to establish the convergence result for the ODE in (37). The proof relies on the optimality properties of the solutions to OPTAVR, Lemma 3 from [30], Theorem 3 (b), and Lemma 6. A detailed proof is given in Appendix F.

Theorem 4: Suppose $\boldsymbol{\theta}^A(\tau)$ evolves according to the ODE in (37). Then, for any initial condition $\boldsymbol{\theta}^A(0) \in \mathcal{H}$, $\lim_{\tau\to\infty} \boldsymbol{\theta}^A(\tau) = \boldsymbol{\theta}^\pi$.

If the Lipschitz hypothesis in Assumptions U.V.1 and U.E. is relaxed, then the conclusions of Lemma 2 hold with continuity replacing Lipschitz continuity. Existence of solutions to the ordinary differential (37) in the set $\mathcal{H}$ follows by Peano's theorem since $\mathcal{H}$ is compact, thus rendering the vector field [associated with (37)] continuous and bounded. Note that Lemma 6 does not require Lipschitz continuity, and nor does the proof of Theorem 4.

C. Convergence of AVR and Proof of Theorem 1

In this subsection, we complete the proof of Theorem 1. We first establish a convergence result for the sequence of iterates of the AVR algorithm $(\boldsymbol{\theta}(t))_{t\in\mathbb{N}}$ based on the associated ODE (37). We do so by viewing (19), (20) as a stochastic approximation update equation, and use a result from [17] that relates the iterates to the ODE (37). We establish the desired convergence result by utilizing the corresponding result obtained for the ODE in Theorem 4. A detailed discussion and proof of the result is given in Appendix G.

Lemma 7: If $\boldsymbol{\theta}(0) \in \mathcal{H}$, then the sequence $(\boldsymbol{\theta}(t))_{t\in\mathbb{N}}$ generated by the Algorithm AVR converges almost surely to $\boldsymbol{\theta}^\pi$.

If we use AVR with a constant step size stochastic approximation algorithm obtained by replacing $1/t$ in (19), (20) with a small positive constant ϵ, we can use results like Theorem 2.2 from Chapter 8 of [17] to obtain a result similar in flavor to that in Lemma 7 (which can then be used to obtain optimality results).

Now we prove Theorem 1 mainly using Lemma 7, and stationarity and ergodicity assumptions.

Proof of Theorem 1: For each $i \in \mathcal{N}$

$$\frac{1}{T}\sum_{t=1}^{T} r_i^*\left(\boldsymbol{\theta}(t), C_t\right) = \frac{1}{T}\sum_{t=1}^{T}\left(r_i^*\left(\boldsymbol{\theta}(t), C_t\right) - r_i^*(\boldsymbol{\theta}^\pi, C_t)\right) \\ + \frac{1}{T}\sum_{t=1}^{T} r_i^*(\boldsymbol{\theta}^\pi, C_t). \tag{42}$$

Using Lemma 7, Lipschitz continuity of $\mathbf{r}^*(., c)$ for any $c \in \mathcal{C}$ [see Lemma 2(a)] and finiteness of $\mathcal{C}$, we can conclude that $|r_i^*(\boldsymbol{\theta}(t), C_t) - r_i^*(\boldsymbol{\theta}^\pi, C_t)|$ converges to 0 a.s. (i.e., for almost all sample paths) as $t \to \infty$. Hence, the first term on right hand side of (42) converges to 0 a.s. as $T \to \infty$. The second term converges to $\mathbb{E}[r_i^*(\boldsymbol{\theta}^\pi, C^\pi)]$ by Birkhoff's Ergodic Theorem (see, for, e.g., [10]). Now, note that $\mathbb{E}[r_i^*(\boldsymbol{\theta}^\pi, C^\pi)] = m_i^\pi$ (see Theorem 3(b) and (35)). Since by Lemma 7, $\lim_{t\to\infty} m_i(t) = m_i^\pi$, part (a) of Theorem 1 is proved.

Next, we prove part (b). Note that for each $i \in \mathcal{N}$

$$\begin{aligned} &\mathrm{Var}^T\left(r_i^*\right) \\ &\quad= \frac{1}{T}\sum_{t=1}^{T}\left(r_i^*\left(\boldsymbol{\theta}(t), C_t\right) - \frac{1}{T}\sum_{s=1}^{T} r_i^*\left(\boldsymbol{\theta}(s), C_s\right)\right)^2 \\ &\quad= \frac{1}{T}\sum_{t=1}^{T}\left(r_i^*\left(\boldsymbol{\theta}(t), C_t\right) - m_i^\pi\right)^2 \\ &\qquad- \left(\frac{1}{T}\sum_{s=1}^{T} r_i^*\left(\boldsymbol{\theta}(s), C_s\right) - m_i^\pi\right)^2. \end{aligned} \tag{43}$$

The second term on the right-hand-side of (43) converges a.s. to zero as $t \to \infty$ by part (a). Also, following the same steps as in the proof of part (a), we see that the first term converges a.s. to v_i^π as $T \to \infty$. Since by Lemma 7, $\lim_{t\to\infty} v_i(t) = v_i^\pi$, part (b) of Theorem 1 is proved.

Part (c) of Theorem 1 follows from parts (a) and (b). ■

VI. Simulations

In this section, we evaluate additional performance characteristics of AVR via simulation. We focus on the realization of different mean-fairness-variability tradeoffs by varying the

functions $(U_i^E, U_i^V)_{i\in\mathcal{N}}$, and on the convergence rate of the algorithm.

For the simulations, we consider a time-slotted setting involving time varying utility functions as discussed in Section II. We consider a network where $N = 20$. Temporal variations in video content get translated into time varying quality rate maps, and we model this as follows: in each time slot, a time varying quality rate map for each user is picked independently and uniformly from a set $\mathcal{Q} = \{q_1, q_2\}$. Motivated by the video distortion versus rate model proposed in [31], we consider the following two (increasing) functions that map video compression rate w to video quality

$$q_1(w) = 100 - \frac{40000}{w - 500}, \quad q_2(w) = 100 - \frac{80000}{w - 500}.$$

These (increasing) functions map video compression rate w to a video quality metric. We see that the map q_2 is associated with a time slot in which the video content (e.g., involving a scene with a lot of detail) is such that it needs higher rates for the same quality (when compared to that for q_1). Referring Section II, we see that $\mathcal{Q}_i = \mathcal{Q}$ for each user $i \in \mathcal{N}$. For each user, the peak data rate in each time slot is modeled as an independent random variable with various distributions (discussed below) from the set $\mathcal{W} = \{\omega_1, \omega_2\}$ where $\omega_1 = 30\,000$ units and $\omega_2 = 60\,000$ units (thus $\mathcal{P} = \mathcal{W}^N$). Further, we choose $r_{\max} = 100$ and the run length of each simulation discussed below is 100 000 time slots.

To obtain different tradeoffs between mean, variability and fairness, for each $i \in \mathcal{N}$ we set $U_i^E(e) = e^{1-\alpha}/(1-\alpha)$ and $U_i^V(v) = \beta\sqrt{v+1}$ and vary α and β. For a given α, note that $U_i^E(\cdot)$ corresponds to α-fair allocation discussed in Section II where a larger α corresponds to a more fair allocation of QoE. Also, by choosing a larger β we can impose a higher penalty on variability. The choice of $U_i^V(\cdot)$ roughly corresponds to the metric proposed in [33]. To obtain *a good initialization* for AVR, the reward allocation in the first 10 time slots is obtained by solving a modified version of $\mathrm{OPTAVR}((\mathbf{m}, \boldsymbol{v}), c)$ with a simpler objective function $\sum_{i\in\mathcal{N}} U_i^E(r_i)$ (which does not rely on any estimates) under the same constraints (17) and (18), and run AVR from the 11th time slot initialized with parameters $(\mathbf{m}, \boldsymbol{v})$ set to the mean reward and half the variance in reward over the first ten time slots.

We first study a homogeneous setting in which, for each time slot, the peak data rate of each user is picked independently and uniformly at random from the set $\mathcal{W}$. Here, we set $\alpha = 1.5$ and vary β over {0.02, 0.1, 0.2, 0.5, 1, 2}. The averaged (across users) values of the mean reward and standard deviation of the reward allocation for the different choices of β are shown in Fig. 1. Not only does the standard deviation reduce with a higher β, we also see that the reduction in mean reward for a given reduction in variability is very small. For instance here we were able to reduce the standard deviation in reward from around 10 to 3 (i.e., around 70% reduction) at the cost of a mere reduction of 4 units in the mean reward (around 7% reduction). It should be clear that the reduction in variance corresponding to the above data will be even more drastic than that of the standard deviation and this is the case in the next setting too.

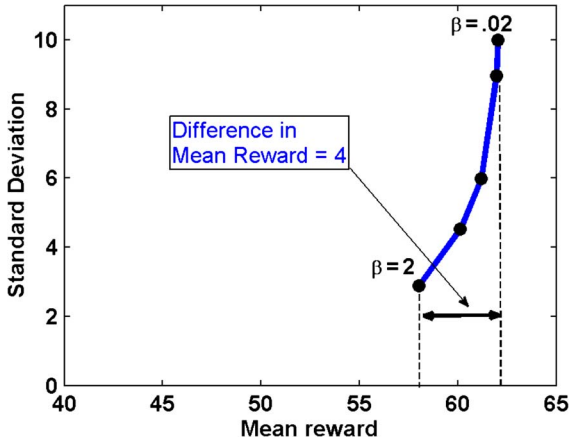

Fig. 1. Homogeneous setting: Mean-variability tradeoffs.

TABLE I
HETEROGENEOUS SETTING: MEAN-VARIABILITY-FAIRNESS TRADEOFFS

α	β	Mean	Variance	Std.Devn.	M_{fair}
0.05	0.02	62.14	65.23	8.08	0.85
0.05	0.1	62.10	49.11	7.01	0.84
0.05	0.5	61.44	19.52	4.42	0.83
0.05	1	60.72	11.66	3.41	0.81
0.05	2	59.25	5.15	2.27	0.79
1.5	0.02	62.06	65.09	8.07	0.89
1.5	0.1	62.00	49.20	7.01	0.89
1.5	0.5	61.37	19.67	4.43	0.88
1.5	1	60.66	11.87	3.44	0.87
1.5	2	59.21	5.23	2.29	0.86
5	0.02	61.86	65.89	8.10	0.93
5	0.1	61.80	49.72	7.05	0.93
5	0.5	61.18	20.122	4.47	0.93
5	1	60.46	11.80	3.44	0.93
5	2	58.87	5.03	2.24	0.92

Next, we study a heterogeneous setting. For each time slot, the peak data rate of each user indexed 1 through 10 is modeled as a random variable taking values ω_1 and ω_2 with probability 0.9 and 0.1 respectively, and that of each user indexed 11 through 20 is ω_1 and ω_2 with probability 0.1 and 0.9, respectively. Thus, in this setting, users with index in the range 1 through 10 typically see poorer channels, and can end up being sidelined if the allocation is not fair. To measure the fairness of a reward allocation, we use a simple metric M_{fairn} which is the ratio of the minimum value to the maximum value of the QoE of the users. In Table I, the value of M_{fair} along with values of the averages (across users) of the mean, variance and standard deviation of the allocated rewards for different choices of α and β are given. As in the homogeneous setting, we see that we can achieve drastic reduction in the variability of quality (measured in terms of either the variance or the standard deviation) for a relatively small reduction in the mean reward. We further see that higher values of α result in a higher values of M_{fair} for the same β, and thus reduce the disparity in allocation of quality.

In Fig. 2, we compare the performance of AVR to that of the optimal offline policy obtained by solving $\mathrm{OPT}(T)$ (for a large T). To enable computation of the optimal offline policy, we consider a setting with just 2 users. We let $U_i^E(e) = e$ and $U_i^V(v) = 0.1v$ for both the users. The time varying quality rate maps are generated just as for the above simulations, and peak

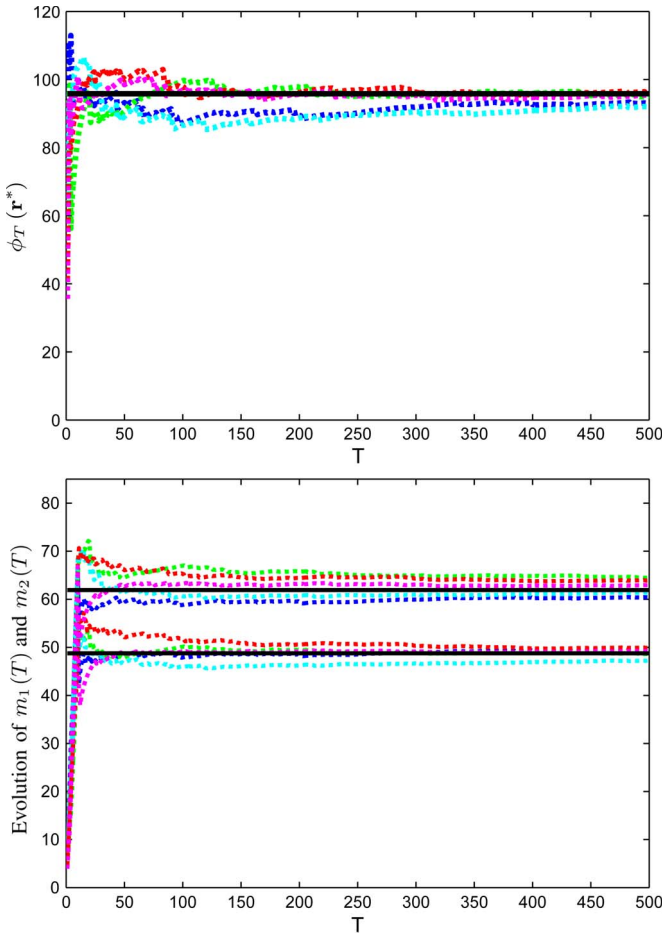

Fig. 2. Performance of AVR (top figure) and convergence of parameters (bottom figure).

data rates are heterogeneous as described above so that the first user sees poorer channels compared to the second. The thick line in the top figure corresponds to the optimal value obtained by solving the offline formulation $\mathrm{OPT}(T)$ for $T = 1500$. The dashed lines depict the performance of AVR in terms of $\phi_t(\mathbf{r}^*)$ for different simulation runs. This indicates that the performance of AVR converges to the (achievable) asymptotic upper bound obtained by solving the offline formulation, i.e., the optimal value of $\mathrm{OPT}(T)$ as T goes to infinity. The dashed lines in the bottom figure depict the evolution of parameters $m_1(.)$ and $m_2(.)$ for the same simulation runs. Here, the lower and upper thick lines correspond to the mean quality of users 1 and 2 respectively corresponding to the solution to $\mathrm{OPT}(T)$ for $T = 1500$.

VII. Conclusion

This work presents an important generalization of NUM framework to account for the deleterious impact of temporal variability allowing for tradeoffs between mean, fairness and variability associated with reward allocations across a set of users. We proposed a simple asymptotically optimal online algorithm AVR to solve problems falling in this framework. We believe such extensions to capture variability in reward allocations can be relevant to a fairly wide variety of systems. Our future work will encompass the possibility of addressing resource allocation in systems with buffering or storage. e.g., energy and/or data storage.

Appendix

A. Proof of Lemma 2

Proof: For $\boldsymbol{\theta} = (\mathbf{m}, \mathbf{v})$, let

$$\Phi_{\boldsymbol{\theta}}(\mathbf{r}) := \sum_{i\in\mathcal{N}} \left(U_i^E\right)'(e_i)\left(r_i - \left(U_i^V\right)'(v_i)(r_i - m_i)^2\right) \quad (44)$$

for $\mathbf{r} \in \mathbb{R}^N$ where $e_i = m_i - U_i^V(v_i)$ for each $i \in \mathcal{N}$. Next, for any $\boldsymbol{\theta}^a, \boldsymbol{\theta}^b \in \mathcal{H}$ and $\mathbf{r} \in [-2r_{\max}, 2r_{\max}]^N$ (any optimal solution to OPTAVR, i.e., minimizer of $\Phi_{\boldsymbol{\theta}}(\mathbf{r})$ subject to constraints is an interior point of this set), let

$$\Delta\Phi(\mathbf{r}, \boldsymbol{\theta}^a, \boldsymbol{\theta}^b) = \Phi_{\boldsymbol{\theta}^b}(\mathbf{r}) - \Phi_{\boldsymbol{\theta}^a}(\mathbf{r}).$$

We prove part (a) (i.e., the Lipschitz continuity with respect to $\boldsymbol{\theta}$ of the optimizer $\mathbf{r}^*(\boldsymbol{\theta}, c)$ of $\Phi_{\boldsymbol{\theta}}(\mathbf{r})$ subject to constraint c) using Proposition 4.32 in [4]. The first condition in the Proposition requires that $\Delta\Phi(\cdot, \boldsymbol{\theta}^a, \boldsymbol{\theta}^b)$ be Lipschitz continuous. To show this, note that for any $\mathbf{r}^c, \mathbf{r}^d \in [-2r_{\max}, 2r_{\max}]^N$

$$\begin{aligned}
&\Delta\Phi(\mathbf{r}^c, \boldsymbol{\theta}^a, \boldsymbol{\theta}^b) - \Delta\Phi(\mathbf{r}^d, \boldsymbol{\theta}^a, \boldsymbol{\theta}^b) \\
&\quad= \sum_{i\in\mathcal{N}} \left(\left(U_i^E\right)'(e_i^a) - \left(U_i^E\right)'(e_i^b)\right)\left(r_i^c - r_i^d\right) \\
&\qquad+ \sum_{i\in\mathcal{N}} \left(U_i^E\right)'(e_i^a)\left(U_i^V\right)'(v_i^a)\left(r_i^d - r_i^c\right)\left(r_i^d + r_i^c - 2m_i^a\right) \\
&\qquad- \sum_{i\in\mathcal{N}} \left(U_i^E\right)'(e_i^b)\left(U_i^V\right)'(v_i^b)\left(r_i^d - r_i^c\right)\left(r_i^d + r_i^c - 2m_i^b\right).
\end{aligned}$$

Using the above expression, Lipschitz continuity and boundedness of $(U_i^{V'})_{i\in\mathcal{N}}$ and $(U_i^{E'})_{i\in\mathcal{N}}$ (see Assumptions U.V.1 and U.E), and boundedness of $\mathbf{r}^a$ and $\mathbf{r}^b$, we can conclude that there exists some positive finite constant η such that

$$\Delta\Phi(\mathbf{r}^c, \boldsymbol{\theta}^a, \boldsymbol{\theta}^b) \le \eta d(\boldsymbol{\theta}^a, \boldsymbol{\theta}^b) d(\mathbf{r}^a, \mathbf{r}^b).$$

Next, we establish the second condition given in the proposition referred to as second order growth condition. For this we use Theorem 6.1 (vi) from [3], and consider the functions L and ψ discussed in the exposition of the theorem. We have

$$L(\mathbf{r}, \boldsymbol{\theta}, \mu, \boldsymbol{\gamma}, c) = \Phi_{\boldsymbol{\theta}}(\mathbf{r}^*) - \Phi_{\boldsymbol{\theta}}(\mathbf{r}) + \mu c(\mathbf{r}) - \sum_{i\in\mathcal{N}} \gamma_i r_i$$

and for $d \in \mathbb{R}^{\mathcal{N}}$, we have

$$\psi_{\mathbf{r}^*}(\boldsymbol{\theta}^a, c)(\mathbf{d}) = \mathbf{d}^{tr} \nabla_{\mathbf{r}}^2 L\left(\mathbf{r}^*(\boldsymbol{\theta}^a, c), \boldsymbol{\theta}^a, \mu^m(c), \boldsymbol{\gamma}^m(c), c\right) \mathbf{d}$$

where $\mu^m(c)$ and $(\gamma_i^m(c) : i \in \mathcal{N})$ are Lagrange multipliers associated with the optimal solution to $\mathrm{OPTAVR}(\boldsymbol{\theta}^a, c)$. Then, using convexity of c we have

$$\psi_{\mathbf{r}^*(\boldsymbol{\theta}^a, c)}(\mathbf{d}) \ge \sum_{i\in\mathcal{N}} 2\left(U_i^E\right)'(e_i^a)\left(U_i^V\right)'(v_i^a)\, d_i^2.$$

Since $(U_i^{V'})_{i\in\mathcal{N}}$ and $(U_i^{E'})_{i\in\mathcal{N}}$ are strictly positive (see Assumptions U.V.1 and U.E), we can conclude that there

exists some positive finite constant η_1 such that $\psi_{\mathbf{r}^*(\boldsymbol{\theta}^a,c)}(\mathbf{d}) \geq \eta_1\|\mathbf{d}\|^2$. Now, using Theorem 6.1 (vi) from [3], we can conclude that second order growth condition is satisfied.

Thus, we have verified the conditions given in Proposition 4.32 in [4], and thus (a) holds. Then, (b) follows from (a) since $\mathcal{C}$ is finite and:

$$E\left[\mathbf{r}^*(\boldsymbol{\theta}, C^\pi)\right] = \sum_{c\in\mathcal{C}} \pi(c)\mathbf{r}^*(\boldsymbol{\theta}, c).$$

■

B. Proof of Theorem 3

Proof: By KKT-OPTSTAT $(\boldsymbol{\rho}_c^\pi : c \in \mathcal{C})$, $(\mu^\pi(c) : c \in \mathcal{C})$ and $((\gamma_i^\pi(c))_{i\in\mathcal{N}} : c \in \mathcal{C})$ satisfy (29)–(31). To show that $\mathbf{r}^*((\mathbf{m}^\pi, \boldsymbol{v}^\pi), c) = \boldsymbol{\rho}_c^\pi$, we verify that $\boldsymbol{\rho}_c^\pi$ satisfies KKT-OPTSTAT$((\mathbf{m}^\pi, \boldsymbol{v}^\pi), c)$. To that end, we can verify that $\boldsymbol{\rho}_c^\pi$ along with $\mu^* = \mu^\pi(c)/\pi(c)$ and $(\gamma_i^* = \gamma_i^\pi(c)/\pi(c) : i \in \mathcal{N})$ satisfy (21)–(23) by using (29)–(31). This proves part (a).

To prove part (b), first note that $(\mathbf{m}^\pi, \boldsymbol{v}^\pi) \in \mathcal{H}^*$ and this follows from (a) and the definitions (see (32), (33)) of $\mathbf{m}^\pi$ and $\boldsymbol{v}^\pi$. Next, note that for any $(\mathbf{m}, \boldsymbol{v}) \in \mathcal{H}^*$ and each $c \in \mathcal{C}$, $\mathbf{r}^*(\mathbf{m}, \boldsymbol{v}, c)$ is an optimal solution to OPTAVR and thus, there exist nonnegative constants $\mu^*(c)$ and $(\gamma_i^*(c) : i \in \mathcal{N})$ such that for all $i \in \mathcal{N}$, and satisfies KKT-OPTAVR given in (21)–(23). Also, since $(\mathbf{m}, \boldsymbol{v}) \in \mathcal{H}^*$, it satisfies (35), (36). Combining these observations, we have that for all $c \in \mathcal{C}$

$$\begin{aligned}\left(U_i^E\right)'\left(E\left[\mathbf{r}^*(\boldsymbol{\theta}, C^\pi)\right] - U_i^V\left(\mathrm{Var}^\pi\left(\mathbf{r}^*(\boldsymbol{\theta}, C^\pi)\right)\right)\right)\\ \Big(r_i^*(\boldsymbol{\theta}, c) - 2\left(U_i^V\right)'\left(\mathrm{Var}^\pi\left(\mathbf{r}^*(\boldsymbol{\theta}, C^\pi)\right)\right)\left(r_i^*(\boldsymbol{\theta}, c)\right.\\ \left. - E\left[\mathbf{r}^*(\boldsymbol{\theta}, C^\pi)\right]\right)\Big) + \gamma_i^* - \mu^*(c)c_i'\left(\mathbf{r}^*(\boldsymbol{\theta}, c)\right) = 0\\ \mu^*(c)c\left(\mathbf{r}^*(\boldsymbol{\theta}, c)\right) = 0\\ \gamma_i^* r_i^*(\boldsymbol{\theta}, c) = 0.\end{aligned}$$

where $\boldsymbol{\theta} = (\mathbf{m}, \boldsymbol{v})$, and $e_i = m_i - U_i^V(v_i)$ for each $i \in \mathcal{N}$. Now for each $c \in \mathcal{C}$, multiply the above equations with $\pi(c)$ and one obtains KKT-OPTSTAT ((29)–(31)) with $(\pi(c)\mu^*(c) : c \in \mathcal{C})$ and $((\pi(c)\gamma_i^*(c))_{i\in\mathcal{N}} : c \in \mathcal{C})$ as associated Lagrange multipliers. From Lemma 3, OPTSTAT satisfies Slater's condition and hence satisfying KKT conditions is sufficient for optimality for OPTSTAT. Thus, we have that $(\mathbf{r}^*(\mathbf{m}, \boldsymbol{v}, c))_{c\in\mathcal{C}}$ is an optimal solution to OPTSTAT. This observation along with uniqueness of solution to OPTSTAT and (35), (36), imply part (b), i.e., $\mathcal{H}^* = \{(\mathbf{m}^\pi, \boldsymbol{v}^\pi)\}$. ■

C. Proof of Lemma 4

Proof: Recall that $\mathcal{H} = [0, r_{\max}]^N \times [0, v_{\max}]^N$ and $v_{\max} = r_{\max}^2$. Note that for any $\boldsymbol{\theta}$ in the interior of $\mathcal{H}$, $z_j(\boldsymbol{\theta}) = 0$ for all j such that $1 \leq j \leq 2N$ from the definition of $C_{\mathcal{H}}(\boldsymbol{\theta})$ and thus we can restrict our attention to the boundary of $\mathcal{H}$. For any $\boldsymbol{\theta}$ on the boundary of $\mathcal{H}$ and $i \in \mathcal{N}$, we can use the facts that $(\bar{\mathbf{g}}(\boldsymbol{\theta}))_i = E[r_i^*(\boldsymbol{\theta}, C^\pi)] - m_i$ and $0 \leq r_i^*(\boldsymbol{\theta}, C^\pi), m_i \leq r_{\max}$, to conclude that $z_i(\boldsymbol{\theta}) = 0$. Similarly, since $v_{\max} = r_{\max}^2$, we can show that $z_j(\boldsymbol{\theta}) = 0$ for any j such that $N + 1 \leq j \leq 2N$. ■

D. Proof of Lemma 5

Proof: For any $\boldsymbol{\theta} \in \widetilde{\mathcal{H}}$, using the definition of $\widetilde{\mathcal{H}}$, we see that $\widetilde{\mathcal{R}}(\boldsymbol{\theta})$ is a non-empty set. For any $c \in \mathcal{C}$, the set $\{\boldsymbol{\rho}_c \in \mathbb{R}^N : c(\boldsymbol{\rho}_c) \leq 0,\ \rho_{c,i} \geq 0\ \forall\ i \in \mathcal{N}\}$ is compact due to continuity (see Assumption C.1) and boundedness (see Assumption C.2) of feasible region associated with functions in $\mathcal{C}$. Thus, $\mathcal{R}_{\mathcal{C}}$ is also compact. Now, note that $\widetilde{\mathcal{R}}(\boldsymbol{\theta})$ is the intersection of a compact set $\mathcal{R}_{\mathcal{C}}$, and Cartesian product of intersection of inverse images of closed sets associated with continuous functions (corresponding to $\mathbb{E}[.]$ and $\mathrm{Var}(\cdot)$) defined over $\mathbb{R}^N$. Thus, $\widetilde{\mathcal{R}}(\boldsymbol{\theta})$ is compact, and this proves (a).

$\widetilde{\mathcal{H}}$ is bounded since $0 \leq m_i \leq r_{\max}$ and $0 \leq v_i \leq r_{\max}^2$ for each $i \in \mathcal{N}$, and each $(\mathbf{m}, \boldsymbol{v}) \in \widetilde{\mathcal{H}}$.

Let $(\overline{\mathbf{m}}, \overline{\boldsymbol{v}})$ be any limit point of $\widetilde{\mathcal{H}}$. Then, there exists a sequence $((\mathbf{m}_n, \boldsymbol{v}_n))_{n\in\mathbb{N}} \subset \widetilde{\mathcal{H}}$, such that $\lim_{n\to\infty}(\mathbf{m}_n, \boldsymbol{v}_n) = (\overline{\mathbf{m}}, \overline{\boldsymbol{v}})$. Let $(\boldsymbol{\rho}_{c,n})_{c\in\mathcal{C}} \in \widetilde{\mathcal{R}}((\mathbf{m}_n, \boldsymbol{v}_n))$ for each $n \in \mathbb{N}$. Since $((\boldsymbol{\rho}_{c,n})_{c\in\mathcal{C}})_{n\in\mathbb{N}}$ is a sequence in the compact set $\mathcal{R}_{\mathcal{C}}$, it has some convergent subsequence $((\boldsymbol{\rho}_{c,n_k})_{c\in\mathcal{C}})_{k\in\mathbb{N}}$. Suppose that the subsequence converges to $(\overline{\boldsymbol{\rho}}_c)_{c\in\mathcal{C}} \in \mathcal{R}_{\mathcal{C}}$. Then

$$\begin{aligned}E[\overline{\rho}_{C^\pi,i}] &= \lim_{k\to\infty} E[\rho_{C^\pi,n_k i}] = \lim_{k\to\infty} m_{n_k i} = \overline{m}_i\\ \mathrm{Var}(\overline{\rho}_{C^\pi,i}) &= \lim_{k\to\infty} \mathrm{Var}(\rho_{C^\pi,n_k i}) \leq \lim_{k\to\infty} v_{n_k i} = \overline{v}_i.\end{aligned}$$

Thus, $(\overline{\boldsymbol{\rho}}_c)_{c\in\mathcal{C}} \in \widetilde{\mathcal{R}}((\overline{\mathbf{m}}, \overline{\boldsymbol{v}}))$, and hence, $(\overline{\mathbf{m}}, \overline{\boldsymbol{v}}) \in \widetilde{\mathcal{H}}$. Thus, $\widetilde{\mathcal{H}}$ contains all its limit points and hence is closed.

To show convexity, consider $(\mathbf{m}_1, \boldsymbol{v}_1), (\mathbf{m}_2, \boldsymbol{v}_2) \in \widetilde{\mathcal{H}}$, and we show that for any given $\alpha \in [0, 1]$, we have $\alpha(\mathbf{m}_1, \boldsymbol{v}_1) + (1-\alpha)(\mathbf{m}_2, \boldsymbol{v}_2) \in \widetilde{\mathcal{H}}$. Let $(\boldsymbol{\rho}_{c,1})_{c\in\mathcal{C}} \in \widetilde{\mathcal{R}}((\mathbf{m}_1, \boldsymbol{v}_1))$ and $(\boldsymbol{\rho}_{c,2})_{c\in\mathcal{C}} \in \widetilde{\mathcal{R}}((\mathbf{m}_2, \boldsymbol{v}_2))$. Hence, $\mathrm{Var}(r_{1i}(C^\pi)) \leq v_{1i}, \mathrm{Var}(r_{2i}(C^\pi)) \leq v_{2i}\ \forall\ i \in \mathcal{N}$. Let $\boldsymbol{\rho}_{c,3} = \alpha\boldsymbol{\rho}_{c,1} + (1-\alpha)\boldsymbol{\rho}_{c,2}$. Thus, for each $i \in \mathcal{N}$

$$E[\rho_{C^\pi,3i}] = \alpha m_1 + (1-\alpha)m_2. \tag{45}$$

Next, note that $\mathrm{Var}(\rho_{C^\pi})$ is a convex function of $(\rho_c)_{c\in\mathcal{C}}$. This can be shown using convexity of square function and linearity of expectation. Thus, for each $i \in \mathcal{N}$

$$\begin{aligned}\mathrm{Var}(\rho_{C^\pi,3i}) &\leq \alpha\mathrm{Var}(\rho_{C^\pi,1i}) + (1-\alpha)\mathrm{Var}(\rho_{C^\pi,2i})\\ &\leq \alpha v_{1i} + (1-\alpha)v_{2i}.\end{aligned} \tag{46}$$

Thus, from (45) and (46), we have that $(\mathbf{r}_3(c))_{c\in\mathcal{C}} \in \widetilde{\mathcal{R}}(\alpha(\mathbf{m}_1, \boldsymbol{v}_1) + (1-\alpha)(\mathbf{m}_2, \boldsymbol{v}_2))$, and thus $\alpha(\mathbf{m}_1, \boldsymbol{v}_1) + (1-\alpha)(\mathbf{m}_2, \boldsymbol{v}_2) \in \widetilde{\mathcal{H}}$. ■

E. Proof of Lemma 6

Proof: The approach used here is similar to that in [30]. Let $\delta > 0$ be given. With $B_\delta(x_0)$ denoting the open ball of radius δ centered at x_0 select $\varepsilon \in (0, \delta)$ such that

$$\max_{\overline{B_\varepsilon(x_0)}} L < \min_{K\setminus B_\delta(x_0)} L. \tag{47}$$

This is possible, since the hypotheses imply that $L(x_0) < L(x)$ for all $x \in K$, $x \neq x_0$. Indeed, consider any solution γ of (40) starting at $x \in K$, with $x \neq x_0$. Then the invariance of K and

(41) imply that the set of ω-limit points of γ is necessarily the singleton $\{x_0\}$. Note that L is non-increasing along trajectories in K and is strictly decreasing along any portion of a trajectory which does not contain x_0. Choose any $t' > 0$ such $\gamma(t) \neq x_0$ for all $t \in [0, t']$ (this is of course possibly by the continuity of $t \mapsto \gamma(t)$). Therefore we must have

$$L(x) = L(\gamma(0)) > L(\gamma(t')) \geq \lim_{t\to\infty} L(\gamma(t)) = L(x_0).$$

Since K is asymptotically stable there exists a decreasing sequence of open sets $\{G_k\}_{k\in\mathbb{N}}$ such that each G_k is invariant with respect to (40) and $\cap_{k\in\mathbb{N}} G_k = K$. By (41)–(47) and the continuity of L and $\nabla L \cdot f$ we can select $n \in \mathbb{N}$ large enough such that

$$\nabla L(x) \cdot f(x) < 0 \qquad \forall x \in \bar{G}_n \setminus B_\varepsilon(x_0) \tag{48a}$$

$$\max_{\overline{B_\varepsilon(x_0)}} L < \min_{\bar{G}_n \setminus B_\delta(x_0)} L. \tag{48b}$$

It is clear by (48a), (48b) that any trajectory starting in $G_n \cap B_\varepsilon(x_0)$ stays in $B_\delta(x_0)$, implying that x_0 is a stable equilibrium. Let γ be any trajectory of (40). Asymptotic stability of K implies that there exists $t_1 > 0$ such that $\gamma(t) \in G_n$ for all $t > t_1$. Also by (48a) there exists $t_2 \geq t_1$ such that $\gamma(t_2) \in G_n \cap B_\delta(x_0)$. Therefore x_0 is asymptotically stable. ■

F. Proof of Theorem 4

Proof: Applying Lemma 3 in [30] and by identifying $\mathcal{V} \equiv \tilde{\mathcal{H}}$, it follows that $\tilde{\mathcal{H}}$ is asymptotically stable for (32). Define

$$L(\boldsymbol{\theta}) = L(\boldsymbol{m}, \boldsymbol{v}) := -\sum_{i\in\mathcal{N}} U_i^E\left(m_i - U_i^V(v_i)\right).$$

Then

$$\begin{aligned}\nabla L(\boldsymbol{\theta}) \cdot \bar{g}(\boldsymbol{\theta}) &= -\sum_{i\in\mathcal{N}} \left(U_i^E\right)'\left(m_i - U_i^V(v_i)\right)\left(\mathbb{E}\left[r_i^*(\boldsymbol{\theta}, C^\pi)\right] - m_i\right. \\ &\quad \left. - \left(U_i^V\right)'(v_i)\left(\mathbb{E}\left[\left(r_i^*(\boldsymbol{\theta}, C^\pi) - m_i\right)^2\right] - v_i\right)\right). \end{aligned} \tag{49}$$

If $\boldsymbol{\theta} \in \tilde{\mathcal{H}}$, then for some $\boldsymbol{\rho} \in \tilde{\mathcal{R}}(\boldsymbol{\theta})$, (49) takes the form

$$\begin{aligned}\nabla L(\boldsymbol{\theta}) \cdot \bar{g}(\boldsymbol{\theta}) &= -\mathbb{E}\left[\Phi_{\boldsymbol{\theta}}\left(\boldsymbol{r}^*(\boldsymbol{\theta}, C^\pi)\right) - \Phi_{\boldsymbol{\theta}}(\boldsymbol{\rho}_{C^\pi})\right] \\ &\quad - \sum_{i\in\mathcal{N}} \left(U_i^E\right)'\left(m_i - U_i^V(v_i)\right)\left(U_i^V\right)'(v_i)\left(v_i - \mathrm{Var}(\rho_{C^\pi,i})\right)\end{aligned} \tag{50}$$

where $\Phi_{\boldsymbol{\theta}}$ is defined in (44). The optimality of $r_i^*(\boldsymbol{\theta}, c)$ for $\mathrm{OPTAVR}((\boldsymbol{m}, \boldsymbol{v}), c)$ and the fact that $\boldsymbol{\rho} \in \tilde{\mathcal{R}}(\boldsymbol{\theta})$ together with Assumptions U.V.1 and U.E. then imply that both terms on the right-hand-side of (50) are nonpositive and that they vanish only if

$$\mathbb{E}\left[r_i^*(\boldsymbol{\theta}, C^\pi)\right] = \mathbb{E}\left[\rho_{C^\pi,i}\right] = m_i \tag{51}$$

$$\mathrm{Var}\left(r_i^*(\boldsymbol{\theta}, C^\pi)\right) = \mathrm{Var}(\rho_{C^\pi,i}) = v_i. \tag{52}$$

In turn, by Theorem 3 these imply that $\boldsymbol{\theta} = \boldsymbol{\theta}^\pi$. Therefore $\nabla L(\boldsymbol{\theta}) \cdot \bar{g}(\boldsymbol{\theta}) < 0$ for all $\boldsymbol{\theta} \in \tilde{\mathcal{H}}$, $\boldsymbol{\theta} \neq \boldsymbol{\theta}^\pi$ and the result follows by Lemmas 4 and 6. ■

G. Proof of Lemma 7

Proof: This proof draws on standard techniques from stochastic approximation (see e.g., [17]). The key idea is to view (19), (20) as a stochastic approximation update equation, and using Theorem 1.1 of Chapter 6 from [17] to relate (19), (20) to the ODE (37). Below, for brevity, we provide details drawing heavily on the framework developed in [17].

In the following, we show that all the Assumptions required to use the theorem are satisfied. The following sets, variables and functions H, $\boldsymbol{\theta}_t$, $\boldsymbol{\xi}_t$, $\mathbf{Y}_t$, ϵ_t, sigma algebras $\mathcal{F}_t$, $\boldsymbol{\beta}_t$, $\delta\mathbf{M}_t$ and the function $\mathbf{g}$ appearing in the exposition of Theorem 1.1 of [17], correspond to the following variables and functions in our problem setting: $H = \mathcal{H}$, $\boldsymbol{\theta}_t = (\mathbf{m}(t), \boldsymbol{v}(t))$, $\xi_t = c_t$, for each $i \in \mathcal{N}$ $(Y_t)_i = r_i^*(t) - m_i(t)$ and $(Y_t)_{i+N} = (r_i^*(t) - m_i(t))^2 - v_i(t)$, $\epsilon_t = 1/t$ for each t, $\mathcal{F}_t$ is such that $(\boldsymbol{\theta}_0, \mathbf{Y}_{i-1}, \xi_i, i \leq t)$ is $\mathcal{F}_t$-measurable, $\boldsymbol{\beta}_t = \mathbf{0}$ and $\delta\mathbf{M}_t = \mathbf{0}$ for each t, $(g((\mathbf{m}, \boldsymbol{v}), c))_i = r_i^*((\mathbf{m}, \boldsymbol{v}), c) - m_i$ and $(g((\mathbf{m}, \boldsymbol{v}), c))_{i+N} = (r_i^*((\mathbf{m}, \boldsymbol{v}), c) - m_i)^2 - v_i$.

Equation (5.1.1) in [17] is satisfied due to our choice of ϵ_t, and (A4.3.1) is satisfied due to our choice of $\mathcal{H}$. Further, (A.1.1) is satisfied as the solutions to OPTAVR are bounded. (A.1.2) holds due to the continuity result in Lemma 2 (a).

We next show that (A.1.3) holds by choosing the function $\bar{\mathbf{g}}$ as follows for each $i \in \mathcal{N}$: $(\bar{g}(\mathbf{m}, \boldsymbol{v}))_i = E[r_i^*((\mathbf{m}, \boldsymbol{v}), C^\pi)] - m_i$, and $(\bar{g}(\mathbf{m}, \boldsymbol{v}))_{i+N} = E[(r_i^*((\mathbf{m}, \boldsymbol{v}), C^\pi) - m_i)^2] - v_i$. Note that the continuity of the function $\bar{\mathbf{g}}$ follows from Lemma 2 (b).

From Section 6.2 of [17], if ϵ_t does not go to zero faster than the order of $1/\sqrt{t}$, for (A.1.3) to hold, we only need to show that the strong law of large numbers holds for $(g(\mathbf{m}, \boldsymbol{v}, C_t))_t$ for any $(\mathbf{m}, \boldsymbol{v})$. The strong law of large numbers holds since $(C_t)_{t\in\mathbb{N}}$ is a stationary ergodic random process and g is a bounded function. Assumptions (A.1.4) and (A.1.5) hold since $\boldsymbol{\beta}_t = \mathbf{0}$ and $\delta\mathbf{M}_t = \mathbf{0}$ for each t. To check (A.1.6) and (A.1.7), we use sufficient conditions discussed in [17] following Theorem 1.1. Assumption (A.1.6) holds since g is bounded. (A.1.7) holds due to the continuity of $g((\mathbf{m}, \boldsymbol{v}), c)$ in $(\mathbf{m}, \boldsymbol{v})$ uniformly in c which follows from the continuity result in Lemma 2 (a), and the finiteness of $\mathcal{C}$. Thus, using Theorem 1.1, we can conclude that on almost all sample paths, $(\boldsymbol{\theta}(t))_{t\in\mathbb{N}}$ converges to some limit set of the ODE (37) in $\mathcal{H}$. From Theorem 4, for any initialization in $\mathcal{H}$, this limit set is the singleton $\{\boldsymbol{\theta}^\pi\}$, and thus the main result follows. ■

REFERENCES

[1] R. Agrawal and V. Subramanian, "Optimality of certain channel aware scheduling policies," in *Proc. Allerton Conf. Commun., Control Comp.*, 2002, [CD-ROM].

[2] D. Bertsekas, *Dynamic Programming and Optimal Control (2 Vol Set)*, 3rd ed. Boston, MA, USA: Athena Scientific, Jan. 2007.

[3] F. Bonnans, J. F. Bonnans, and A. D. Ioffe, Quadratic Growth and Stability in Convex Programming Problems With Multiple Solutions 1995.

[4] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*. New York, NY, USA: Springer, 2000.

[5] A. Bouch and M. A. Sasse, "Network quality of service: What do users need?" in *Proc. 4th Int. Distrib. Conf.*, 1999, pp. 21–23.
[6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[7] Y. Cai, T. Wolf, and W. Gong, "Delaying transmissions in data communication networks to improve transport-layer performance," *IEEE J. Selected Areas Commun.*, vol. 29, no. 5, pp. 916–927, May 2011.
[8] J. A. Filar, L. C. M. Kallenberg, and H. M. Lee, "Variance-penalized Markov decision processes," *Math. Oper. Res.*, vol. 14, no. 1, pp. 147–161, 1989.
[9] F. X. Frei, R. Kalakota, A. J. Leone, and L. M. Marx, "Process variation as a determinant of bank performance: Evidence from the retail banking study," *Manag. Sci.*, vol. 45, no. 9, pp. 1210–1220, Sep. 1999.
[10] R. M. Gray, *Probability, Random Processes, Ergodic Properties*, 2nd ed. New York, NY, USA: Springer, 2009.
[11] C. Hess, R. Seri, and C. Choirat, "Ergodic theorems for extended real-valued random variables," *Stoch. Processes Appl.*, vol. 120, no. 10, pp. 1908–1919, 2010.
[12] V. Joseph and G. de Veciana, Jointly Optimizing Multi-User Rate Adaptation for Video Transport Over Wireless Systems: Mean-Fairness-Variability Tradeoffs, Technical Report, Jul. 2011. [Online]. Available: www.ece.utexas.edu/~gustavo/VariabilityAwareVideoRateAdapt.pdf
[13] V. Joseph and G. de Veciana, "Variability aware network utility maximization," in *Proc. CoRR*, 2011, [CD-ROM].
[14] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res.*, vol. 49, no. 3, pp. 237–252, 1998.
[15] T. Kim and M. Ammar, "Optimal quality adaptation for scalable encoded video," *IEEE J. Selected Areas Commun.*, vol. 23, no. 2, pp. 344–356, Dec. 2005.
[16] H. Kushner and P. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1250–1259, Jul. 2004.
[17] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY, USA: Springer, 2003.
[18] M. Lin, A. Wierman, L. Andrew, and E. Thereska, "Online dynamic capacity provisioning in data centers," in *Proc. Allerton Conf. Commun., Control Comp.*, 2011, [CD-ROM].
[19] S. Mannor and J. N. Tsitsiklis, "Mean-variance optimization in Markov decision processes," in *Proc. CoRR*, 2011, [CD-ROM].
[20] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.
[21] M. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 2, pp. 396–409, Apr. 2008.
[22] L. Real, J. Ott, and E. Silverfine, "On the tradeoff between the mean and the variance in foraging: Effect of spatial distribution and color preference," *Ecology*, vol. 63, no. 6, pp. 1617–1623, 1982.
[23] P. Seeling, M. Reisslein, and B. Kulapala, "Network performance evaluation using frame size and quality traces of single-layer and two-layer video: A tutorial," *IEEE Commun. Surveys Tutorials*, vol. 6, no. 3, pp. 58–78, 2004.
[24] S. Sen, J. Rexford, J. Dey, J. Kurose, and D. Towsley, "Online smoothing of variable-bit-rate streaming video," *IEEE Trans. Multimedia*, vol. 2, no. 1, pp. 37–48, Mar. 2000.
[25] S. Shakkottai and R. Srikant, "Network optimization and control," *Found. Trends Netw.*, vol. 2, no. 3, pp. 271–379, Jan. 2007.
[26] M. J. Sobel, "Mean-variance tradeoffs in an undiscounted mdp," *Math. Oper. Res.*, vol. 42, no. 1, pp. 175–183, 1994.
[27] J. C. Spall, *Introduction to Stochastic Search and Optimization*. New York, NY, USA: Wiley, 2003.
[28] R. A. Steiner and M. F. Walter, "The effect of allocation schedules on the performance of irrigation systems with different levels of spatial diversity and temporal variability," *Agricultural Water Manag.*, vol. 23, no. 3, pp. 213–224, 1993.
[29] A. L. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Queueing Syst. Theory Appl.*, vol. 50, pp. 401–457, Aug. 2005.
[30] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *Oper. Res.*, vol. 53, pp. 12–25, Jan. 2005.
[31] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Selected Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, Jun. 2000.
[32] N. Tsikriktsis and J. Heineke, "The impact of process variation on customer dissatisfaction: Evidence from the U.S. domestic airline industry," *Decision Sciences*, vol. 35, no. 1, pp. 129–141, 2004.
[33] C. Yim and A. C. Bovik, "Evaluation of temporal variation of video quality in packet loss networks," *Image Commun.*, vol. 26, no. 1, pp. 24–38, Jan. 2011.

Vinay Joseph received the B.Tech. degree in electronics and communication engineering from the National Institute of Technology Calicut, Kerala, India, in 2007, the M.E. degree in telecommunication engineering from the Indian Institute of Science, Bangalore, in 2009, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin in 2013.

He is currently with Qualcomm Corp. R&D, San Diego, CA. His research interests include communication networks, optimization and stochastic approximation.

Gustavo de Veciana (S'88–M'94–SM'01–F'09) received the B.S., M.S, and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 1987, 1990, and 1993, respectively.

He joined the Department of Electrical and Computer Engineering where he currently is the Joe. J. King Professor. He served as the Director and Associate Director of the Wireless Networking and Communications Group (WNCG) at the University of Texas at Austin, from 2003 to 2007. His research focuses on the analysis and design of wireless and wireline telecommunication networks; architectures and protocols to support sensing and pervasive computing; applied probability and queueing theory.

Dr. de Veciana received the National Science Foundation CAREER Award 1996 and five best paper awards including: IEEE William McCalla Best ICCAD Paper Award for 2000, Best Paper in ACM TODAES Jan 2002–2004, Best Paper in ITC 2010, Best Paper in ACM MSWIM 2010, and Best Paper IEEE INFOCOM 2014. He served as Editor and is currently serving as editor-at-large for the IEEE/ACM TRANSACTIONS ON NETWORKING. He is on the Technical Advisory Board of IMDEA Networks.

Ari Arapostathis (F'07) received the B.S. degree from MIT, Cambridge, MA, USA and the Ph.D. degree from the University of California at Berkeley.

He is with the University of Texas at Austin, where he is a Professor in the Department of Electrical and Computer Engineering. His research interests include stochastic optimization and control, adaptive control theory, and hybrid systems.