

# Enhancing Mobile Video Streaming by Lookahead Rate Allocation in Wireless Networks

Hatem Abou-zeid\*, Hossam S. Hassanein\*, and Nizar Zorba†

\*Queen's University, Canada, h.abouzeid@queensu.ca, hossam@cs.queensu.ca

†Qatar Mobility Innovations Center, P.O.Box 210531, Doha, Qatar, nizarz@qmic.com

**Abstract**—Developing novel video delivery mechanisms have become imperative to cope with the unprecedented growth in mobile video traffic. In this paper, we present video transmission schemes that improve the streaming experience by looking *ahead* at the *future* rates users are expected to face. Such an approach is useful for the delivery of stored videos that can be strategically buffered in advance at the users' devices. For instance, if it is known a user is entering a low coverage area, content can be prebuffered to support smooth streaming. Therefore, the Base Stations (BSs) can now plan long-term multi-user rate allocations based not only on current channel states, but also on future conditions. To provide a performance benchmark we first develop a *lookahead* multi-objective Linear Program (LP) that offers a trade-off between minimizing overall network video degradation, and providing fairness in individual user degradation. Then, to efficiently solve the problem, we present a polynomial-time algorithm that closely follows the pareto-optimal trade-off of the multi-objective LP. We provide an extensive performance analysis of the proposed methods by simulations, and numerical results demonstrate that significant improvements in video streaming are achievable by the lookahead rate allocation strategies.

## I. INTRODUCTION

As mobile traffic continues to grow at unprecedented rates, network operators are faced with a myriad of challenges to cope with this demand. Recent forecasts project a 16-fold growth of mobile traffic from 2012 to 2017, with video accounting for 66 percent of the total traffic [1]. Furthermore, in vehicular environments, users experience rapid channel fluctuations which result in streaming disruptions as users traverse the network. Consequently, research and standardization efforts are focusing on devising novel video delivery mechanisms.

In this paper, we present video transmission schemes that improve the streaming experience by looking *ahead* at the future rates users are anticipated to receive. Our focus is on optimizing the delivery of stored videos that can be buffered in advance at the user devices. Being aware of a user's upcoming rate allows the network to plan rate allocations that prevent or reduce video degradations. For instance, if a user is moving towards the cell edge or a tunnel, the network can increase the allocated wireless resources allowing the user to buffer more video content. *Pre-buffering* this additional data then provides smooth video streaming since the user can consume the buffer while being in poor radio coverage. If, on the other hand,

the user is approaching the BS, transmission can be delayed, provided sufficient video content has previously been buffered. Therefore, by coupling knowledge of the users' buffer status and future data rates, the BS can devise *lookahead* rate allocation strategies that enhance the streaming experience. We refer to such schemes as Lookahead Rate Allocation (LRA) for video streaming.

The proposed LRA schemes are based on exploiting wireless channel predictions, which are generally possible due to the correlation between location and the achievable rate [2]. Therefore, if a user's future location is known, the upcoming data rates can be anticipated from radio and coverage maps stored at the network, which can be updated in real-time from User Equipment (UE) measurements [3], [4]. While such predictions are particularly plausible for users in public transportation, trains, or vehicles on highways, studies on human mobility patterns reveal a high degree of temporal and spatial regularity, suggesting a potential 93% average predictability [5]. Furthermore, a key motivation for incorporating such predictions is the plethora of navigation and context information available in today's smart phones which can facilitate lookahead allocation mechanisms.

### A. Related Work and Contributions

Using rate predictions to enhance mobile video streaming has been discussed in a limited number of recent works. In [6], Yao et al. develop a rate adaptation algorithm that proactively switches to the predicted transmission rates by consulting a stored radio map. Unlike in our paper, the authors do not intend to pre-buffer content based on predictions but to improve TCP rate control and throughput by faster convergence to the available capacity. Similar to our approach, the work in [7] also proposes prebuffering video content based on a bandwidth look-up service which is constantly updated by users traversing the network. This is achieved by allowing users headed to poor conditions to request additional segments in advance. However, the presented mechanisms are based on a distributed user-centric solution and are limited to a single user. Since they do not address the multi-user resource allocation problem, this prevents obtaining network wide objectives or efficiently trading-off video quality among multiple users.

The work in references [8] and [9] is closer to our work. Algorithms that employ rate predictions to minimize system utilization and avoid streaming delays are proposed in [8]. However, the authors focus on single-cell, low load scenarios

This work was made possible by a *National Priorities Research Program* (NPRP) grant NPRP 09-406-2-153 from the Qatar National Research Fund (Member of Qatar Foundation).

TABLE I  
 SUMMARY OF IMPORTANT SYMBOLS

Symbol	Description
$i$	User index, $i = \{1, 2, \dots, M\}$
$k$	BS index, $k = \{1, 2, \dots, K\}$
$\mathcal{K}$	Set of BSs in the network
$\mathcal{M}$	Set of users in the network
$n$	Time slot index, $n = \{1, 2, \dots, N\}$
$N$	Number of slots in the lookahead window
$\mathcal{N}$	Set of time slots in the lookahead window
$\hat{r}_{i,n}$	Link rate of user $i$ at slot $n$ [bits]
$\tau$	Duration of a time slot [s]
$\mathcal{U}_{k,n}$	Set containing the indices of users associated with BS $k$ at slot $n$
$x_{i,n}$	Fraction of air-time assigned to user $i$ at slot $n$

where system utilization can be reduced, and also do not consider fairness among users. In our own work [9], a multi-user, multi-cell rate allocation LP is solved to minimize streaming degradations over multiple users. However, due to the large problem dimension, the LP can only be solved offline, and is therefore only suitable for benchmarking. Further, the presented formulation does not incorporate any measures to provide fairness among the users in the streaming quality. In this paper, we extend the work in [9], with the following main contributions:

- We develop a LRA scheme that exploits rate predictions to enhance video streaming, while enabling a *trade-off* between overall network streaming quality and *fairness* in individual user quality. The problem is formulated as a multi-objective LP which provides a benchmark solution.
- To efficiently solve the aforementioned LRA problem, we develop a polynomial-time algorithm. Results indicate that the proposed heuristic performs close to the LP benchmark, at a fraction of the memory and computation requirements. Further, the algorithm is tunable and follows the pareto-optimal trade-off of the multi-objective LP.

## B. Organization of the paper

The rest of this paper is organized as follows. First, in Section II we outline the system model, assumptions, and notation. The multi-objective LP formulation of Lookahead Rate Allocation (LRA) for stored video streaming is developed in Section III, while Section IV presents the proposed near-optimal algorithm. In Section V, we evaluate the resulting video streaming enhancements of the proposed schemes under several network settings. Finally, a discussion of future directions and implementation issues of LRA is made in Section VI.

## II. SYSTEM MODEL

We use the following notational conventions:  $\mathcal{X}$  denotes a set and its cardinality  $|\mathcal{X}|$  is denoted by  $X$ . We use bold letters to denote matrices, e.g.  $\mathbf{x} = (x_{a,b} : a \in \mathbb{Z}_+, b \in \mathbb{Z}_+)$ , and  $(x)^+$  denotes  $\max\{0, x\}$ . Frequently used notation is summarized in Table I.

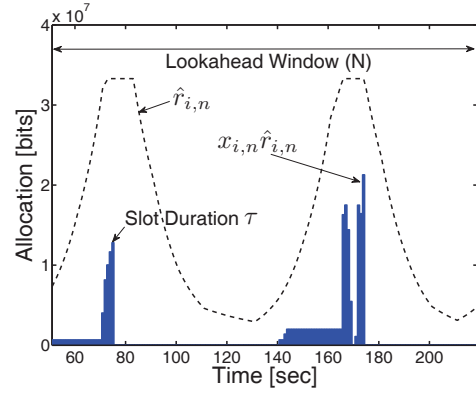


Fig. 1. Sample user allocation during the lookahead window of  $N$  time slots.

## A. Network Overview

Consider a network with a BS set  $\mathcal{K}$  and an active user set  $\mathcal{M}$ . An arbitrary BS is denoted by  $k \in \mathcal{K}$  and a user by  $i \in \mathcal{M}$ . Users request stored video content that is transported using an HTTP-based progressive download mechanism. We assume that the wireless link is the bottleneck, and therefore the requested video content is always available at the BS for transmission.

## B. Link Model and Resource Sharing

Time is divided in slots of equal duration  $\tau$ , and a slot is denoted by  $n \in \mathcal{N}$ , where the set of considered time slots  $\mathcal{N} = \{1, 2, \dots, N\}$ . In each slot, the wireless channel can be shared among multiple users during which the achievable data rate is assumed to be constant for each user<sup>1</sup>. A typical value of such a coherence time  $\tau$  is 1 s for vehicle speeds up to 20 m/s, during which average wireless capacity is not significantly affected. The achievable data rate depends on the path loss model  $PL(d) = 128.1 + 37.6 \log_{10} d$ , where the user-BS distance  $d$  is in km [10]. The feasible link rate is computed using Shannon's equation with SNR clipping at 20 dB for practical modulation orders. Therefore, a user  $i$  at slot  $n$ , will have a feasible data transmission of

$$r_{i,n} = \tau B \log_2(1 + P_{rx,i,n}/N_o B) \quad [\text{bits}] \quad (1)$$

where  $P_{rx}$ ,  $N_o$  and  $B$  are the received power, noise power spectral density, and the transmission bandwidth respectively.

User link rates are assumed to be known for the upcoming  $N$  slots, which we call the *lookahead window*. A matrix of future link rates is defined by  $\hat{\mathbf{r}} = (\hat{r}_{i,n} : i \in \mathcal{M}, n \in \mathcal{N})$ . Fig. 1 illustrates an example of  $\hat{r}_{i,n}, \forall n$  for a user traversing two BSs along a highway. In this paper, we assume that knowledge of  $\hat{\mathbf{r}}$  is error free to provide the bounds of the potential gains.

BS air-time is shared among the active users during each slot  $n$ . We define the resource allocation matrix  $\mathbf{x} = (x_{i,n} \in [0, 1] : i \in \mathcal{M}, n \in \mathcal{N})$  which gives the fraction of time during each slot  $n$  that the BS bandwidth is assigned to user  $i$ . The

<sup>1</sup>As we are interested in long-term allocation planning, we only consider average rate variations.

rate received by each user at each slot is the element-wise product  $\mathbf{x} \odot \hat{\mathbf{r}}$ . A sample allocation  $x_{i,n}, \forall n$  for a user  $i$  is illustrated in Fig. 1, where the bars indicate the proportion of  $\hat{r}_{i,n}$  allocated to that user. Note that since a user can traverse multiple cells during  $N$ , BS cooperation is needed to make the allocation plan. This is assumed to be possible via an inter-BS interface such as the X2-interface in Long Term Evolution (LTE) networks.

User-BS association is based on the strongest received signal, and is assumed to be known during  $N$ . We define the set  $\mathcal{U}_{k,n}, k \in \mathcal{K}, n \in \mathcal{N}$  which contains the indices of all the users associated with BS  $k$  at time slot  $n$ .

### III. LOOKAHEAD RATE ALLOCATION FOR STREAMING: PROBLEM FORMULATION

The objective of LRA is to determine the rate allocation matrix  $\mathbf{x}$  that minimizes the video streaming degradations that users experience. This is achieved by exploiting the rate prediction matrix  $\hat{\mathbf{r}}$  such that users opportunistically prebuffer video content during their peak rates, before poor channel conditions prevail. To do so, we first define the following video degradation metric [9].

#### A. Video Degradation (VD) Metric

This metric captures the difference between the cumulative number of bits the user is requesting at the streaming bitrate, and the cumulative rate allocated to the user. It is computed for each slot, and therefore a matrix of values  $\text{VD}_{i,n}$  is obtained. If the streaming bitrate is  $V$  [bps], the cumulative requirement at slot  $n$  is  $V\tau n$  bits. Since the cumulative number of bits allocated to a user by slot  $n$  is  $\sum_{n'=1}^n x_{i,n'} \hat{r}_{i,n'}$ , therefore video degradation is defined as

$$\text{VD}_{i,n} = [V\tau n - \sum_{n'=1}^n x_{i,n'} \hat{r}_{i,n'}]^+. \quad (2)$$

When  $\sum_{n'=1}^n x_{i,n'} \hat{r}_{i,n'} > V\tau n$  it means more is allocated than required, indicating that future video content is pre-buffered, and  $\text{VD}_{i,n} = 0$ . On the other hand, if the converse holds, then the user will experience video stalling or quality degradation, which we refer to as video degradation. Therefore, VD represents the amount of unfulfilled video demand. The average network VD over  $N$  slots is therefore:

$$\text{VD}_{\text{Net}} = \frac{1}{NM} \sum_{n=1}^N \sum_{i=1}^M \text{VD}_{i,n}. \quad (3)$$

We can also compute the *future* VD that a user at slot  $n$  will experience (denoted by  $\tilde{\text{VD}}_{i,n}$ ) based on the current cumulative allocation at slot  $n$ , and a tentative air-time allocation for the slots  $n+1, n+2, \dots, N$ . This is obtained as follows

$$\tilde{\text{VD}}_{i,n} = \sum_{n'=n}^N [V\tau n' - \sum_{n''=1}^{n'} x_{i,n''} \hat{r}_{i,n''}]^+, \quad (4)$$

where  $n'$  and  $n''$  are dummy variables. A high value of  $\tilde{\text{VD}}_{i,n}$  indicates that the user does not have content pre-buffered, and that the tentative future allocation is insufficient

(possibly because the user is headed towards poor channel conditions/congested zones). This measure will be used by the algorithm in Section IV to prioritize such users and pre-buffer their video content before the poor conditions prevail.

#### B. Pareto-Optimal Video Degradation Minimization

The objective of the pareto-optimal VD minimization is to jointly minimize the *total* network VD and the *individual* user video degradations during the  $N$  slots. This can be achieved by the following multi-objective optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \alpha \sum_{i=1}^M \sum_{n=1}^N \frac{\text{VD}_{i,n}}{\tau n V} + \beta \max_i \sum_{n=1}^N \frac{\text{VD}_{i,n}}{\tau n V} \quad (5) \\ \text{subject to:} \quad & \text{C1: } \sum_{i \in \mathcal{U}_{k,n}} x_{i,n} \leq 1, \quad \forall k, n \\ & \text{C2: } \sum_{n=1}^N \hat{r}_{i,n} x_{i,n} \leq \tau n V, \quad \forall i \\ & \text{C3: } 0 \leq x_{i,n} \leq 1 \quad \forall i, n. \end{aligned}$$

The first term in the objective is the normalized network VD, while the second term represents a min max objective for the normalized individual user VD, and  $\alpha, \beta \in [0, 1]$ . Constraint C1 expresses the resource limitation at each base station. It ensures that the sum of the air-time of all users associated with BS  $k$  is equal to 1 at every time slot. C2 limits the amount of video content delivered to a user during the  $N$  slots to the total amount request by that user. Finally, C3 provides the bounds for the resource allocation factor. Although the constraints in (5) are linear, the objective is not, due to the  $[\cdot]^+$  operator in (2) and the min max component of the objective function. However, by introducing auxiliary variables we can express (5) as the following equivalent LP:

$$\begin{aligned} \underset{\mathbf{x}, \mathbf{D}, Y}{\text{minimize}} \quad & \alpha \sum_{i=1}^M \sum_{n=1}^N \frac{D_{i,n}}{\tau n V} + \beta Y \quad (6) \\ \text{subject to:} \quad & \text{C1, C2, C3} \\ \forall i, n \quad & \text{C4: } \tau n V - \sum_{n'=1}^n x_{i,n'} \hat{r}_{i,n'} - D_{i,n} \leq 0, \\ \forall i \quad & \text{C5: } - \sum_{n=1}^N \frac{D_{i,n}}{\tau n V} + Y \leq 0, \\ \forall i, n \quad & \text{C6: } D_{i,n} \geq 0. \end{aligned}$$

In this reformulation, by introducing deficiency  $D_{i,n}$  as additional optimization variables and restricting them to have positive values (in C6), we have removed the  $[\cdot]^+$  operator. Minimizing  $D_{i,n}$  is now equivalent to minimizing  $\text{VD}_{i,n}$  as computed in (2). Similarly, the min max objective component can be replaced with the variable  $Y$  and C5. Constraint C5 ensures that  $Y$  takes the value of the largest user VD, and therefore minimizing it is equivalent to minimizing the max of user VD. Although the problem is now linear, it has a large number of constraints  $KN + 2M + MN$  and  $2MN + M$  optimization variables. This can be solved with

large-scale LP solvers such as Gurobi [11] but will require significant memory and considerable time to solve for large problem sizes. Therefore, (6) can only serve as an offline performance benchmark, which we refer to as VDMIN-Opt. For real-time implementation, we present a heuristic algorithm in the following section.

#### IV. PROPOSED LRA ALGORITHM

The main idea of the proposed VD minimization algorithm is to first keep track of the *cumulative* rates allocated to users up to slot  $n$ . Then, the *future* video degradation users are predicted to experience are computed, and resource allocations  $x_{i,n}$  are made such that the user VDs are minimized.

##### A. Algorithm Steps

As in the VDMIN-Opt formulation, this algorithm jointly minimizes  $VD_{Net}$  and individual user VD. The essence of the algorithm is the definition of a new rate allocation *metric* to make air-time allocations  $x_{i,n}$ . The metric can also be tuned to tradeoff fairness in user VD with network  $VD_{Net}$ . In more detail, it consists of the following steps:

- *Step 1:* Initialize  $x_{i,n} = 0$  for all the users and time slots.
- *Step 2:* Compute the *future*  $\tilde{VD}_i$  each user will experience at slot  $n$ , using (4).
- *Step 3:* Each base station  $k$  allocates the full air-time at slot  $n$  to the user  $i^*$  (i.e.  $x_{i^*,n} = 1$ ) that satisfies

$$i^* = \arg \max_i r_{i,n} \tilde{VD}_{i,n}^\gamma \quad \forall i \in \mathcal{U}_{k,n}. \quad (7)$$

The intuition of this allocation metric is to prioritize users with *both* a high *current* channel quality and a high *future* video degradation. Therefore, users will opportunistically pre-buffer their content when their channel quality is good, and before poor future conditions prevail. The parameter  $\gamma$  controls the influence of the future user VD in the metric. A higher  $\gamma$  will prioritize users with VD and provide more fairness.

- *Step 4:* Repeat steps 2 and 3 for all  $n \in \mathcal{N}$ .
- *Step 5:* Calculate  $VD_{Net}$  using (3).
- *Step 6:* Repeat steps 2-5 until there is no more decrease in  $VD_{Net}$ .

Note in the first iteration,  $\mathbf{x} = 0$  in the computation of (4), and therefore step 3 will not exploit future VD information. However, subsequent iterations of steps 2-5 will allocate  $x_{i,n}$  based on the values of  $x_{i,n'} \forall n' = n+1, n+2, \dots, N$  of the previous iteration. As  $\hat{r}_{i,n}$  does not change over iterations, the selection in step 3, changes in the direction of decreasing VD. Typically, the algorithm converges within 4 to 6 iterations, as observed for the various network and mobility settings in Section V. The complete procedure is presented in Algorithm 1, which we refer to as VDMIN-Alg.

##### B. Computational Complexity

The computational complexity of VDMIN-Alg is primarily dominated by the computation of  $\tilde{VD}_{i,n}$  in (4) which takes  $O(N^2)$  time for each user. Therefore, the overall complexity is of the order  $O(MN^3)$  to make the allocation plan for the upcoming  $N$  slots in a network with  $M$  users.

---

#### Algorithm 1 LRA Video Streaming Algorithm: VDMIN-Alg

---

**Require:**  $\hat{r}_{i,n}, \mathcal{U}_{k,n}, V, \tau, M, K, N$

- 1: Initialize  $x_{i,n}, R_{i,n} = 0 \quad \forall i, n$
  - 2: **repeat** {allocation iterations}
  - 3:   Calculate  $VD_{Net}$  using (3).
  - 4:   **for all** time slots  $n$  **do**
  - 5:     Reset  $x_{i,n} = 0 \quad \forall i$ .
  - 6:     **for all** base stations  $k$  **do**
  - 7:       **for all** users  $i \in \mathcal{U}_{k,n}$  **do**
  - 8:         Calculate the *future* VD using (4)
  - 9:       **end for**
  - 10:       Set  $x_{i^*,n} = 1$  to  $i^*$  with the highest  $\hat{r}_{i,n} \tilde{VD}_{i,n}^\gamma$ .
  - 11:     **end for**
  - 12:   **end for**
  - 13:   Calculate  $VD_{Net}$  after allocation.
  - 14: **until** {no more decrease in  $VD_{Net}$ }
  - 15: **return**  $\mathbf{x}$
- 

#### V. SIMULATION RESULTS

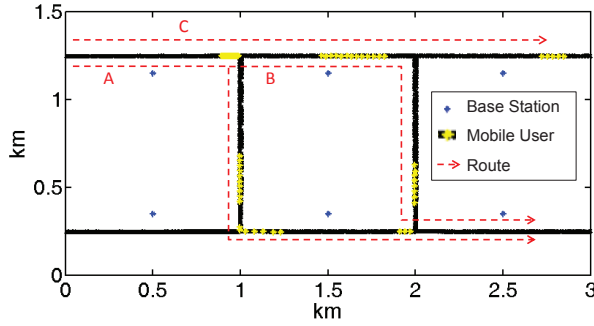
##### A. Simulation Set-up

To provide general results we consider two network and mobility scenarios. The first is the six BS network shown in Fig. 2(a), that covers the illustrated road network. Realistic vehicular mobility is generated on the roads using the SUMO traffic simulator [12]. Vehicles traverse the three routes denoted by A, B and C in Fig. 2(a) with equal probability. For evaluation in a more general (but less practical) network, we also model the 19 cell network illustrated in Fig. 2(b), where users move according to the Random Way Point (RWP) mobility model at a constant speed  $S$ , zero pause time between the waypoints, and no wrap-around. This model enables the study of LRA with users experiencing independent sequences of data rate fluctuations. The inter-BS distance is set to 1 km, with a BS transmit power of 40 W, a center carrier frequency of 2 GHz, and a bandwidth of 10 MHz. The user speed  $S$  is set to 10 m/s for the RWP mobility model, and the video streaming rate  $V$  is set to 3 Mbits. We consider a lookahead window  $N$  of 200 slots with a slot duration  $\tau$  of one second. Simulations are repeated 50 times to obtain the average values of following metrics.

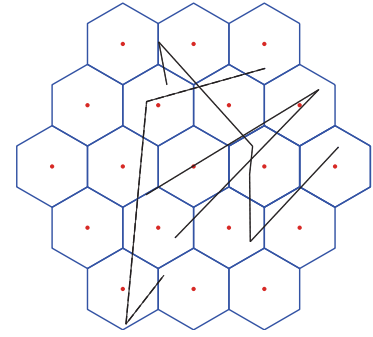
- $VD_{Net}$ : the average network VD as defined in (3).
- $J_{Net}^{VD}$ : Jain's fairness index for user VD during  $N$ , and is computed as  $(\sum_{i=1}^M VD_i)^2 / M \sum_{i=1}^M VD_i^2$ , where  $VD_i$  is the average individual user VD during the lookahead window  $N$ .

We compare the performance of the LRA schemes against two baseline approaches that do not exploit rate predictions: Equal Share (ES) and Rate-Proportional (RP). In ES, air-time is shared equally among the users at each time slot. If there are  $N_{k,n}$  users associated with BS  $k$  at time  $n$ , then  $x_{i,n} = 1/N_{k,n} \forall i \in \mathcal{U}_{k,n}$ . The RP allocator is designed to be more spectrally efficient but not completely fair to users. Here, the air-time assigned to each user  $i$  at slot  $n$





(a) Road network with the considered mobility routes.



(b) The 19 cell network with RWP mobility.

Fig. 2. Network and mobility models considered.

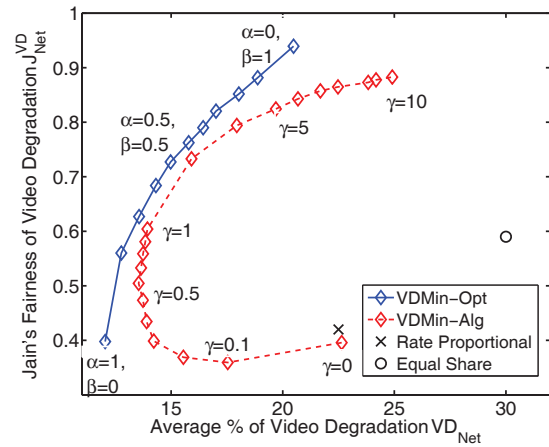
is in proportion to the achievable data-rate of that user  $\hat{r}_{i,n}$ . Therefore,  $x_{i,n} = \hat{r}_{i,n} / \sum_{i \in \mathcal{U}_{k,n}} \hat{r}_{i,n}$  in RP.

### B. Performance Evaluation

Fig. 3 illustrates the pareto-optimal trade-off between  $VD_{\text{Net}}$  and  $J_{\text{Net}}^{\text{VD}}$  that the LRA mechanisms achieve. First, we observe the significant improvements in *both*  $VD_{\text{Net}}$  and VD fairness of the proposed lookahead schemes compared to the RP and ES baseline allocators. We can see that video degradation can be reduced by 50% without sacrificing fairness. The figure also demonstrates how the proposed VDMin-Alg closely follows the pareto-optimal benchmark curve of the VDMin-Opt that is solved offline. By varying  $\gamma$  in VDMin-Alg, the  $VD_{\text{Net}} - J_{\text{Net}}^{\text{VD}}$  trade-off can be effectively controlled. Further, the deviation from the benchmark VDMin-Opt is lower in the region that offers a good trade-off, with  $\gamma = 1$  providing a suitable operating point.

In Fig. 4(a) we show the effect of varying the number of users on the experienced video degradation for the RWP scenario. To illustrate the lowest achievable VD we set  $\alpha = 1, \beta = 0$  and refer to this setting as VDMin-Opt-Greedy. We can clearly see the large VD savings which increase with network load compared to the base-line RP and ES allocators. While the baseline RP outperforms the ES it has a lower fairness in VD as illustrated in Fig. 4(b). VDMin-Opt-Greedy also performs poorly in terms of VD fairness, particularly at higher loads. In Fig. 4(a) and Fig. 4(b) we also depict the performance of the proposed LRA algorithm (VDMin-Alg) at the selected operating point of  $\gamma = 1$ . We can see that it closely follows the benchmark VDMin-Opt-Greedy in reducing VD, while simultaneously providing significant fairness gains.

In Fig. 5 a similar study is conducted on the road network of Fig. 2(a). We observe that VDMin-Alg follows VDMin-Opt even more closely in terms of VD minimization, thereby supporting its generality. The fairness gains are less in this case, which is expected, as all the users follow similar trajectories and therefore all the allocators provide a reasonable degree of fairness. Nevertheless, VDMin-Alg still provides some fairness gains at close to optimal VD minimization.


 Fig. 3. Pareto-optimal trade-off in VD; RWP scenario with  $M = 100$ .

## VI. CONCLUSION

In this paper, we presented lookahead rate allocation strategies that can significantly enhance mobile video streaming. To provide a benchmark, we first formulated a multi-objective LP that captures the required trade-off between minimizing total video degradation and achieving fairness in video quality among users. Then, we presented an LRA algorithm that closely follows the benchmark solution in polynomial time, and with a significantly lower memory requirement. Numerical results indicate that the LRA approaches consistently achieve significant improvements in the video streaming quality over traditional allocation schemes, for different user mobility patterns and network loads.

There are several directions for future work. First of all, studies on the effects of prediction errors are needed to assess performance under uncertainty. This includes errors in the rate predictions of the coverage maps, and in the predicted user trajectories. Furthermore, models that capture the variability and accuracy of the predicted information, as well as methods and algorithms that are robust to such inaccuracies are needed. Research on developing decentralized lookahead rate allocation algorithms operating at each BS independently

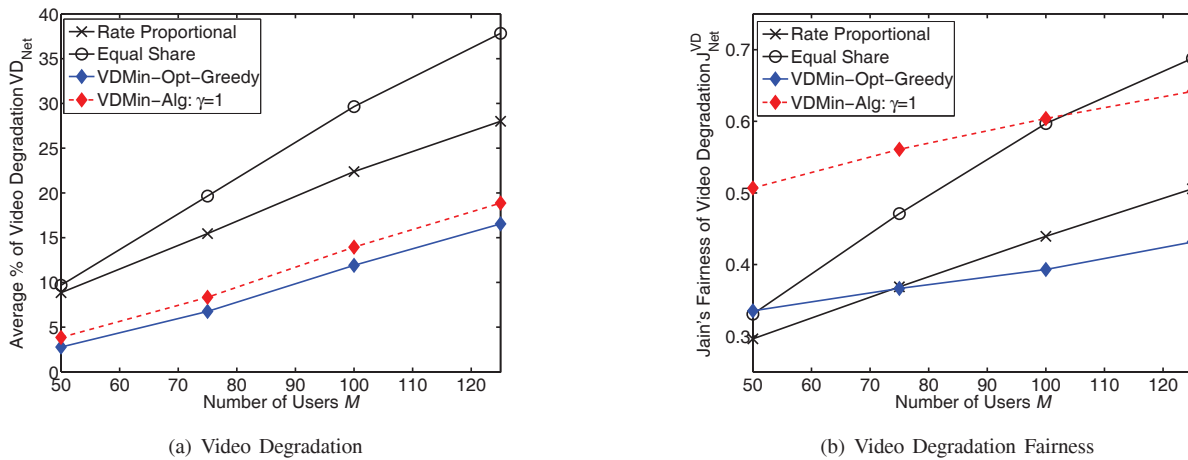


Fig. 4. Network Video Degradation and Jain's Fairness of Video Degradation vs. number of users with RWP mobility.

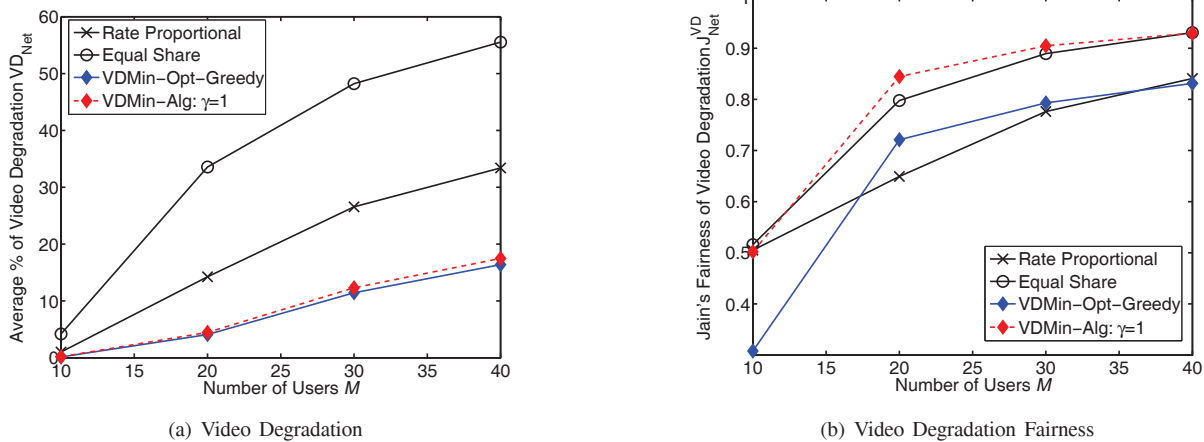


Fig. 5. Network Video Degradation and Jain's Fairness of Video Degradation vs. number of users with road network mobility.

is another potential direction of work. In [13] we provide a more detailed discussion on the required signalling, network functional elements and their interaction, in order to develop predictive/lookahead wireless access mechanisms. Therefore, while the results demonstrate that exploiting rate predictions in lookahead allocation strategies provides valuable video streaming quality gains, further work is needed to address the aforementioned research issues.

REFERENCES

[1] CISCO, "Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017," 2013.  
 [2] M. Malmirchegini and Y. Mostofi, "On the spatial predictability of communication channels," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 964-978, Mar. 2012.  
 [3] J. Johansson, W. Hapsari, S. Kelley, and G. Bodog, "Minimization of drive tests in 3GPP release 11," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 36-43, 2012.  
 [4] OpenSignal, "The OpenSignal project homepage." <http://opensignal.com/>, 2013. Link verified on Feb. 15th, 2013.  
 [5] C. Song, Z. Qu, N. Blumm, and A. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, pp. 1018-1021, Feb. 2010.

[6] J. Yao, S. Kanhere, and M. Hassan, "Improving QoS in high-speed mobility using bandwidth maps," *IEEE Trans. Mobile Comput.*, vol. 11, pp. 603-617, Apr. 2012.  
 [7] H. Riiser, T. Endestad, P. Vigmostad, C. Griwodz, and P. Halvorsen, "Video streaming using a location-based bandwidth-lookup service for bitrate planning," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 8, pp. 24:1-24:19, Aug. 2012.  
 [8] Z. Lu and G. de Veciana, "Optimizing stored video delivery for mobile networks: The value of knowing the future," in *Proc. IEEE Int. Conf. on Computer Commun. (INFOCOM)*, pp. 2806-2814, July 2013.  
 [9] H. Abou-zeid, H. S. Hassanein, and S. Valentin, "Optimal predictive resource allocation: Exploiting mobility patterns and radio maps," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, to appear.  
 [10] 3GPP, "LTE/E-UTRA; radio frequency system scenarios," Technical Report TR 36.942 V11.0.0, 3GPP, Sept. 2012.  
 [11] Gurobi Optimization. <http://www.gurobi.com/>. Link verified on Jul. 23th, 2013.  
 [12] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "SUMO-simulation of urban mobility: An overview," in *Proc. Third Int. Conf. on Advances in System Simulation (SIMUL 2011)*, pp. 63-68, Oct. 2011.  
 [13] H. Abou-zeid and H. S. Hassanein, "Predictive green wireless access: Exploiting mobility and application information," *IEEE Commun. Mag.*, vol. 20, Oct. 2013, to appear.