



CHALMERS

Chalmers Publication Library

On Proactive Caching with Demand and Channel Uncertainties

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

53rd Annual Allerton Conference on Communication, Control, and Computing

Citation for the published paper:

Muppirisetty, L. ; Tadrous, J. ; Eryilmaz, A. et al. (2015) "On Proactive Caching with Demand and Channel Uncertainties". 53rd Annual Allerton Conference on Communication, Control, and Computing

Downloaded from: <http://publications.lib.chalmers.se/publication/228134>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

On Proactive Caching with Demand and Channel Uncertainties

L. Srikar Muppirisetty¹, John Tadrous², Atilla Eryilmaz³, and Henk Wymeersch¹

Abstract—Mobile data traffic has surpassed that of voice to become the main component of the system load of today’s wireless networks. Recent studies indicate that the data demand patterns of mobile users are predictable. Moreover, the channel quality of mobile users along their navigation paths is predictable by exploiting their location information. This work aims at fusing the statistically predictable demand and channel patterns in devising proactive caching strategies that alleviate network congestion. Specifically, we establish a fundamental bound on the minimum possible cost achievable by any proactive scheduler under time-invariant demand and channel statistics as a function of their prediction uncertainties, and develop an asymptotically optimal proactive service policy that attains this bound as the prediction window grows. In addition, the established bound yields insights on how the demand and channel statistics affect proactive caching decisions. We reveal some of these insights through numerical investigations.

I. INTRODUCTION

The major bottleneck for future networks lies in handling the increasingly high mobile data traffic from the users. It is predicted that mobile data traffic will see a nine-fold increase by the end of 2020 [1]. Therefore, networks should consider employing various smart resource allocation strategies to comply with this rapidly rising user data demands. Interestingly, the demand patterns of the mobile users [2]–[4] and also their experienced channel quality metrics (CQM) including, e.g., received signal strength, RMS delay spread, and interference levels are predictable [5]–[8].

Today’s networks predominantly employ reactive strategies for resource allocation in which the user requests are served by the network after being initiated by the user. This reactive approach suffers from huge penalty when there is heavy traffic from the users requesting data. In contrast, proactive resource allocation strategies, which are seen as one of the key disruptive technologies for 5G wireless networks [9], can track, learn, and then predict the user demand requests ahead of time, and hence possess more flexibility in scheduling these requests before their actual time of arrival. The main advantage of this approach is network load balancing over large time scale dynamics, at the expense of possible waste of network resources [10].

*This research was supported, in part, by the European Research Council, under Grant No. 258418 (COOPNET); by STINT Grant IB2013-5301; by NSF Grants: CAREER-CNS-0953515, CNS-WiFiUS-1456806, and CCSS-EARS-1444026; and the DTRA Grant: HDTRA1-15-1-0003. Also, the work of A. Eryilmaz was supported, in part, by the QNRF Grant: NPRP 7-923-2-344.

¹L. Srikar Muppirisetty and Henk Wymeersch are with Chalmers University of Technology, Sweden. E-mail: {srikar.muppirisetty, henkw}@chalmers.se.

²John Tadrous is with Rice University, USA. E-mail: jtadrous@rice.edu.

³Atilla Eryilmaz is with The Ohio State University, USA. E-mail: eryilmaz.2@osu.edu.

Proactive design strategies have gained importance very recently. The work in [11] introduced a novel proactive resource allocation paradigm by exploiting the predictability of the user behavior. It provided a solid theoretical background and demonstrated significant spectral efficiency gains in various scenarios. In [10], proactive resource allocation schemes under time-invariant and time-varying demand statistics are studied. The authors proposed fundamental lower bounds on the achievable costs, and developed asymptotically optimal policies that approach these bounds when the prediction window size is increased. In [12], [13], the impact of proactive scheduling on user delay has been investigated under idealistic queuing scenarios whereby it is shown that proactive scheduling reduces average delay exponentially in proactive window size. However, the utilization of predicted CQM for proactive resource allocation is not captured in proactive scheduling thus far. Energy efficient scheduling such as in [14], [15] studied optimal policies with statistically known CQM, but assumed perfectly predictable demand. Clearly, there is a gap in the research area of designing optimal proactive schedulers with demand and channel uncertainties. In [5]–[8], a location-aided framework was proposed and showed how large-scale channel characteristics of the wireless channel can be predicted by exploiting the user’s location information. Since location-aided predicted CQM is coarse, it can be efficiently harnessed in predictive/proactive resource allocation whereby demand dynamics and large scale channel characteristics vary within the same time scale.

In this work, we study proactive resource allocation strategies that exploit both the predictable data demand and channel characteristics, with uncertainties. Especially, we consider a time-invariant demand statistics model in which all the incoming requests from the user over time are statistically indistinguishable. We further consider time-invariant channel statistics model in which user experiencing a channel state over time is independent and identically distributed. The main contributions of this paper can be summarized as follows:

- We extend the work in [10] with the inclusion of predicted channel statistics of the users and study its effect on the proactive scheduling.
- We establish a global lower bound on the proactive scheduling cost that captures the impact of demand and channel uncertainties.
- Driven by insights from the obtained bound, we develop a stationary and asymptotically optimal proactive service policy that even approaches that bound as prediction window size grows.

- We demonstrate that the designed proactive scheduler offers better performance in terms of lower achievable cost in contrast to the baseline reactive scheduler.

Notation: Vectors and matrices are written in bold (e.g., a vector \mathbf{k} and a matrix \mathbf{K}); $\mathbb{E}[\cdot]$ denotes expectation; $\{\cdot\}_t$ denotes a collection of elements, and $(\cdot)_t$ denotes a sequence of elements.

II. SYSTEM MODEL

We consider a network which consists of a set of N users $\mathcal{N} = \{1, 2, \dots, N\}$. The network is equipped with a service provider which serves the user data requests by allocating resources to them on a per time slot basis.

User demand model

We assume the data requests from the users is not perfectly predictable but with uncertainties. Let $d_{n,t} \in \{0, 1\}$ be a binary random variable that captures the probabilistic nature of the data request from a user n in time slot t , where $d_{n,t} = 1$ indicates a data request with probability $\pi_{n,t} = P(d_{n,t} = 1)$. Each user $n \in \mathcal{N}$ over an infinite time period generates independent sequence of data requests $\{d_{n,t}\}_t$. We assume time-invariant demand statistics for the data requests from the users¹. Then, $\{d_{n,t}\}_t$ for each user n is an independent and identically distributed sequence of random variables with $\mathbb{E}[d_{n,t}] = \bar{\pi}_n$. Let us denote the demand profile of the users as $\bar{\pi} = (\bar{\pi}_n)_n$. We assume that the service provider is aware of the demand profile $\bar{\pi}$, which captures the statistical characteristics of future demand. Each request from the user is served by the network by spending S amount of its resources².

Wireless channel model

We assume each user n in a time slot t experiences one of the K_n possible channel states from the set $\mathcal{C}_n = \{g_n^{(k)}, k = 1, \dots, K_n\}$ with corresponding probabilities $\psi_n = \{\psi_n^{(k)}, k = 1, \dots, K_n\}$, such that $\sum_{k=1}^{K_n} \psi_n^{(k)} = 1, \forall n \in \mathcal{N}$. It should be noted that the set of channel states \mathcal{C}_n , the number of possible channel states K_n , and the possible channel states $g_n^{(k)}$ are all dependent on user index n . Therefore, different users tend to have different support sets of channel realizations. All channel realizations are assumed to be non-negative and finite. We further assume time-invariant channel statistics for the channel states of the users across different time slots. In other words, user n experiences a channel state $g_n^{(k)}$, which is independent and identically distributed across different time slots. The channel profile of users is denoted as $\Psi = (g_n^{(k)}, \psi_n^{(k)})_{n=1, \dots, N}$.

Let us define a channel realization vector $\mathbf{g} = [g_1, g_2, \dots, g_n]$ denoting a possible realization of channels for all n users. The probability of this vector being realized

¹The system can further be generalized to time-varying (fluctuating) demand characteristics as in [10], yet this will lead to complicated notation without significant conceptual benefit. Hence, we have not considered this scenario here.

²The results obtained in this work can directly be generalized to the case where such amount of resources is user and time-dependent, i.e., $S_{n,t}$, yet known to the system within the prediction window of T slots in advance.

is $P_c(\mathbf{g})$. Note that $\mathbf{g} \in \mathcal{C}$, where $\mathcal{C} = \prod_n \mathcal{C}_n$. We can thus write $P_c(\mathbf{g}) = \prod_n P(g_n)$. We further denote $\{\mathbf{g}_t\}_t$ as a sequence of channel realization of all users over time t .

Cost function model

We denote $C_d(x)$, $C_d : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, as the demand cost function for serving a demand $x \geq 0$ in a time slot. The demand cost function $C_d(x)$ is assumed to be strictly convex and increasing. We denote $C_c(g)$, $C_c : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, as the channel cost function to utilize the channel $g \geq 0$ in a time slot. The channel cost function $C_c(g)$ is decreasing³ in g .

Reactive network model

We consider reactive network as a baseline scenario in which the requests are served upon their arrival. The load of a user n in time slot t for a reactive network under channel realization $g_{n,t}$ is written as

$$L_{n,t}^{\mathcal{R}}(g_{n,t}) = S d_{n,t} C_c(g_{n,t}). \quad (1)$$

The time-averaged expected cost of all users under reactive operation is

$$c^{\mathcal{R}}(\bar{\pi}, \Psi) = \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \mathbb{E} \left[C_d \left(\sum_{n=1}^N L_{n,t}^{\mathcal{R}}(g_{n,t}) \right) \right], \quad (2)$$

where expectation is over the demand and channel statistics of the users.

Proactive network model

We assume the service provider is aware of the demand $\bar{\pi}$ and channel Ψ profiles of the users over a T time slot prediction window. As shown in [10], the network traffic load is averaged out over time if the service provider employs a proactive operation. Following the same notation as in [10], let us denote $u_{n,t}(\tau)$ as the amount of proactive service applied to a user n at time slot t for a future possible request after τ slots⁴, i.e., at time $t + \tau$, where $1 \leq \tau \leq T$. The proactive service of a future request cannot exceed the total demand of S units of service, i.e.,

$$\sum_{\tau=1}^T u_{n,t-\tau}(\tau) \leq S, \forall n, t, \quad (3)$$

and the proactive service can never be negative, i.e.,

$$u_{n,t}(\tau) \geq 0, \forall n, t, \tau. \quad (4)$$

Let us denote the amount of load generated by a user n in a time slot t as $L_{n,t}^{\mathcal{P}}(\mathbf{u}_{n,t}, g_{n,t})$ under proactive control

³We can view S , for example, as the total number of bits $B = \sum_t b_t$ to be delivered over a T time slot period. There is a channel cost for sending b_t bits in each time slot. The channel cost is related to the amount of energy spent E_t to send b_t bits over the wireless channel g . Obviously, the cost is more if channel g is bad and vice-versa. So, the cost E_t is inversely related to the channel state g .

⁴The notation of the proactive service $u_{n,t}(\tau)$ can best understood with an example. Consider the case with $t = 1$ and $\tau = 2$, then $u_{n,1}(2)$ indicates the proactive service applied in time slot 1 for a future possible request in time slot 3, i.e., two slots ahead of the current time slot.

$\mathbf{u}_{n,t} = (u_{n,t}(\tau))_\tau$ and the channel realization $g_{n,t}$, which is written as

$$L_{n,t}^{\mathcal{P}}(\mathbf{u}_{n,t}, g_{n,t}) = \left(S - \sum_{\tau=1}^T u_{n,t-\tau}(\tau) \right) d_{n,t} C_c(g_{n,t}) + \sum_{\tau=1}^T u_{n,t}(\tau) C_c(g_{n,t}), \quad (5)$$

where the term $\sum_{\tau=1}^T u_{n,t-\tau}(\tau)$ corresponds to the past applied proactive services for each user n and the term $\sum_{\tau=1}^T u_{n,t}(\tau)$ captures the proactive service to be applied for user n over the next T slots.

Problem statement

The goal of the proactive controller is to determine the optimal online proactive service policy that minimizes the time averaged expected cost while delivering the content on time. The optimization problem of the proactive controller is written as $c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) =$

$$\begin{aligned} \min_{\{u_{n,t}(\tau)\}_{n,t,\tau}} \quad & \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \mathbb{E} \left[C_d \left(\sum_{n=1}^N L_{n,l}^{\mathcal{P}}(\mathbf{u}_{n,l}, g_{n,l}) \right) \right] \\ \text{s.t.} \quad & (3), (4). \end{aligned} \quad (6)$$

The subscript T captures the proactive service window size, and the superscript \mathcal{P} indicates proactive operation.

III. LOWER BOUND AND ASYMPTOTICALLY OPTIMAL DESIGN

Under the time-invariant demand statistics model, the incoming demand requests from the user over time are statistically indistinguishable. Under the time-invariant channel statistics model, a channel state experienced by a user is independent and identically distributed across different time slots. For this model, the proactive gains come from the ability to harness available demand and channel statistics together with observable instantaneous realizations of demand and channels to flatten the network load over time at minimum cost. As in [10], the exact solution of (6) is not tractable due to infinite dimensionality of the problem. Instead, we resort to find a global lower bound on the achievable cost as a function of demand and channel uncertainties, and develop an asymptotically optimal policy which attain that bound as proactive window size grows.

A. Global lower bound on minimum achievable cost

In this subsection, we state a fundamental, non-trivial bound on achievable costs by any proactive policy.

Theorem 1: Let $\mathcal{B}_t = \{n \in \mathcal{N} : d_{n,t} = 1\}$ be the set of users that generate data requests at time t according to $\bar{\boldsymbol{\pi}}$, and $\mathbf{g}_t \in \mathcal{C}$ be the channel realization vector of all the users at time t according to $\boldsymbol{\Psi}$. Then, under time-invariant demand and channel statistics model and for any $T \geq 1$, the optimal proactive scheduling cost $c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi})$ of (6), satisfies

$$c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) \geq \mathcal{C}_{\mathcal{U}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) \quad (7)$$

where

$$\begin{aligned} \mathcal{C}_{\mathcal{U}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) = \min_{\{\tilde{\mu}_n(\mathcal{B}, \mathbf{g})\}_{n, \mathbf{g}, \mathcal{B}}} & \left\{ \sum_{\mathbf{g} \in \mathcal{C}} P_c(\mathbf{g}) \sum_{\mathcal{B} \subseteq \mathcal{N}} P_d(\mathcal{B}) \times \right. \\ & C_d \left(\sum_{n \in \mathcal{B}} \left(S C_c(g_n) - \left(\sum_{\mathbf{h} \in \mathcal{C}} P_c(\mathbf{h}) \right. \right. \right. \\ & \times \sum_{\mathcal{D} \subseteq \mathcal{N}} P_d(\mathcal{D}) \tilde{\mu}_n(\mathcal{D}, \mathbf{h}) \left. \left. \left. \right) C_c(g_n) \right) \right) \\ & \left. + \sum_{n=1}^N \tilde{\mu}_n(\mathcal{B}, \mathbf{g}) C_c(g_n) \right\} \\ & \text{subject to } 0 \leq \tilde{\mu}_n(\mathcal{B}, \mathbf{g}) \leq S, \forall n, \mathcal{B}, \mathbf{g} \quad (8) \end{aligned}$$

where $P_d(\mathcal{B})$ is the probability of set $\mathcal{B}_t = \mathcal{B}$ under the time-invariant demand statistics model, $P_c(\mathbf{g})$ is the probability of channel realization \mathbf{g} under the time-invariant channel statistics model.

Proof: See Appendix I. ■

In the objective of (8), the term $\sum_{\mathbf{h} \in \mathcal{C}} P_c(\mathbf{h}) \sum_{\mathcal{D} \subseteq \mathcal{N}} P_d(\mathcal{D}) \tilde{\mu}_n(\mathcal{D}, \mathbf{h})$ corresponds to the average proactive service assigned to a request from user n before the request is actually realized. The term $\tilde{\mu}_n(\mathcal{B}, \mathbf{g})$ is the total expected proactive service assigned to all possible requests from user n when the current set of demanding users is \mathcal{B} and their channel realization is \mathbf{g} . We note that the optimization of $\mathcal{C}_{\mathcal{U}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi})$ is convex and yields a unique solution due to strict convexity of $C_d(\cdot)$. The theorem establishes that no proactive scheduling policy can achieve a lower cost than the non-trivial bound $\mathcal{C}_{\mathcal{U}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi})$ under the uncertainties associated with the time-invariant demand and channel statistics. In contrast to [10, Theorem 1] with only time-invariant demand statistics model, the lower bound $\mathcal{C}_{\mathcal{U}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi})$ accounts for the additional information of channel statistics available at the scheduler.

B. Asymptotically optimal design

In this subsection, we develop a simple stationary policy that hinges on the solution of (8) to deliver efficient performance in moderate T regimes, and asymptotically realizes the lower bound $\mathcal{C}_{\mathcal{U}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi})$.

Definition 1 (Policy \mathbf{p}): Let $\{\mu_n(\mathcal{B}, \mathbf{g})\}_{n, \mathcal{B}, \mathbf{g}}$ denote the optimal solution to (8). We consider proactive scheduling policy \mathbf{p} that observes the set of requesting users \mathcal{B}_t , and channel gain realizations \mathbf{g}_t every time slot t , and assigns proactive controls $u_{n,t}(\tau) = \frac{\mu_n(\mathcal{B}_t, \mathbf{g}_t)}{T}$, $\forall n, t, 1 \leq \tau \leq T$.

From the definition, the online policy \mathbf{p} is stationary and depends only on the current realization of demand and channel quality. In addition, it is a simple policy that directly employs a look-up table of length $2^N \prod_n K_n$ which entails a search process of complexity $O(N + \sum_n \log(K_n))$. Note that, to apply policy \mathbf{p} , the solution of (8) has to be obtained offline based on the demand $\bar{\boldsymbol{\pi}}$ and channel $\boldsymbol{\Psi}$ profiles.

Now, we establish the asymptotic optimality property of policy \mathbf{p} .

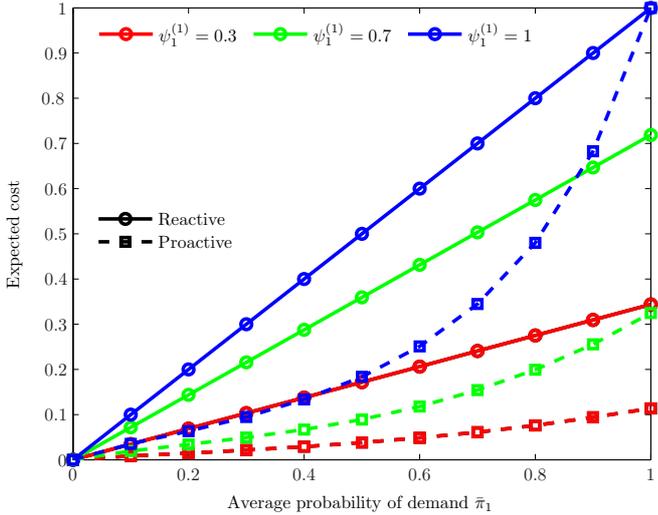


Fig. 1. Comparison of reactive and proactive costs for single user scenario under the time-invariant demand and channel statistics model. The probabilities of the demand $\bar{\pi}_1$ and channel $\psi_1^{(1)}$ are varied with the channel states fixed to $g_1^{(1)} = 1$ (bad channel state), $g_1^{(2)} = 2$ (good channel state).

Theorem 2: Denote the time average expected cost under policy \mathbf{p} by $c_T^{\mathbf{p}}(\bar{\pi}, \Psi)$. Then policy \mathbf{p} is asymptotically optimal, in the sense that $\limsup_{T \rightarrow \infty} |c_T^{\mathbf{p}}(\bar{\pi}, \Psi) - c_T^{\mathbf{p}^*}(\bar{\pi}, \Psi)| = 0$.

Proof: Please refer to Appendix II. ■

Thanks to strong law of large numbers, equal allocation of proactive service throughout the prediction window of size T , policy \mathbf{p} achieves the global lower bound as $T \rightarrow \infty$. Having established the key characteristics of proactive scheduling under demand and channel uncertainties, we next move on to deeper insights on the system performance through numerical simulations.

IV. NUMERICAL RESULTS AND DISCUSSION

A. Scenario

The network scheduler is aware of the user demand $\bar{\pi}$ and channel Ψ profiles. The scheduler spends $S = 1$ units of service for each request. Each user n observes one of the two possible channel states $\{g_n^{(1)}, g_n^{(2)}\}$ with probabilities $\{\psi_n^{(1)}, 1 - \psi_n^{(1)}\}$. We consider $g_n^{(2)} \geq g_n^{(1)}$, hence $g_n^{(2)}$ is termed the good channel state, while $g_n^{(1)}$ is the bad channel state. We assume the cost function for the demand to be of polynomial form $C_d(x) = x^4$ and for the channel usage as $C_c(g) = 1/g$.

B. Impact of demand and channel probabilities on the expected cost

We consider a single user scenario to understand the impact of demand and channel statistics on expected cost for reactive and proactive schedulers. The time averaged expected cost of the reactive scheme can be computed from (2) as

$$\begin{aligned} c^{\mathcal{R}}(\bar{\pi}_1, g_1^{(1)}, g_1^{(2)}, \psi_1^{(1)}) \\ = \bar{\pi}_1 \left((1/g_1^{(1)})^4 \psi_1^{(1)} + (1/g_1^{(2)})^4 (1 - \psi_1^{(1)}) \right). \end{aligned} \quad (9)$$

We denote $\mu_1 = \tilde{\mu}_1(0, g_1^{(1)})$, $\mu_2 = \tilde{\mu}_1(1, g_1^{(1)})$, $\mu_3 = \tilde{\mu}_1(0, g_1^{(2)})$, $\mu_4 = \tilde{\mu}_1(1, g_1^{(2)})$ to be make the notation simpler. The proactive controls μ_1 and μ_3 corresponds to the two channels states when there is no demand requests, and the corresponding controls with demand requests are μ_2 and μ_4 . Then the lower bound to the proactive scheduling cost can be computed from (8) as

$$\begin{aligned} \underline{c}_{\mathcal{U}}(\bar{\pi}_1, g_1^{(1)}, g_1^{(2)}, \psi_1^{(1)}) \\ = \min_{(\mu_1, \mu_2, \mu_3, \mu_4) \geq 0} \psi_1^{(1)} (1 - \bar{\pi}_1) \mu_1^4 (1/g_1^{(1)})^4 + \psi_1^{(1)} \bar{\pi}_1 \\ \times (1/g_1^{(1)})^4 \left(1 - \psi_1^{(1)} (1 - \bar{\pi}_1) \mu_1 - \psi_1^{(1)} \bar{\pi}_1 \mu_2 - (1 - \psi_1^{(1)}) \right. \\ \times (1 - \bar{\pi}_1) \mu_3 - (1 - \psi_1^{(1)}) \bar{\pi}_1 \mu_4 + \mu_2 \left. \right)^4 + (1 - \psi_1^{(1)}) \\ \times (1 - \bar{\pi}_1) \mu_3^4 (1/g_1^{(2)})^4 + (1 - \psi_1^{(1)}) \bar{\pi}_1 (1/g_1^{(2)})^4 \\ \times \left(1 - \psi_1^{(1)} (1 - \bar{\pi}_1) \mu_1 - \psi_1^{(1)} \bar{\pi}_1 \mu_2 - (1 - \psi_1^{(1)}) \right. \\ \times (1 - \bar{\pi}_1) \mu_3 - (1 - \psi_1^{(1)}) \bar{\pi}_1 \mu_4 + \mu_4 \left. \right)^4 \end{aligned} \quad (10)$$

Fig. 1 compares the average costs $c^{\mathcal{R}}(\bar{\pi}_1, g_1^{(1)}, g_1^{(2)}, \psi_1^{(1)})$ and $\underline{c}_{\mathcal{U}}(\bar{\pi}_1, g_1^{(1)}, g_1^{(2)}, \psi_1^{(1)})$, achievable by the reactive and proactive schemes respectively. It can be observed that the proactive scheduling offers lower cost compared to the reactive scheme for any $(\bar{\pi}_1, \psi_1^{(1)})$. The reactive scheme bears no freedom in the scheduling strategy to minimize the cost as it has to serve the demand requests after they have been initiated, whereas the proactive scheme offers flexibility in the scheduling strategy by exploiting the demand and channel statistics to minimize cost by load balancing. The expected cost for both schemes increases with increase in $\bar{\pi}_1$ and $\psi_1^{(1)}$. The cost increases with increase in $\bar{\pi}_1$ due to the fact that the system is more loaded with incoming demand requests. The reason for the cost to increase with $\psi_1^{(1)}$ is as the user more often experiences a bad channel state $g_1^{(1)}$ than a good channel state $g_1^{(2)}$.

It is interesting to note that, unlike in the case with time-invariant demand statistics and no channel knowledge, the cost of reactive and proactive services do not converge when $\bar{\pi}_1 = 1$ (which was reported in [10]). The reason for this behavior can be best understood from (10) by setting $\bar{\pi}_1 = 1$. It can be derived easily that proactive service coincides with the reactive service when the user always observes either the good or bad channel state all the time. Hence, when the channels are identical over all slots and data demand is certain, then there is no gain of applying proactive service. However, when the channels vary from one slot to another (i.e., $0 < \psi_1^{(1)} < 1$), then even with certain data demand there is still potential to apply proactive service in the presence of good channel so as to minimize the cost when the bad channel is realized.

C. Impact of the value of channel states on the expected cost

Fig. 2 compares the impact of increase in $g_1^{(2)}$ on the cost of reactive and proactive schemes for a single user scenario.

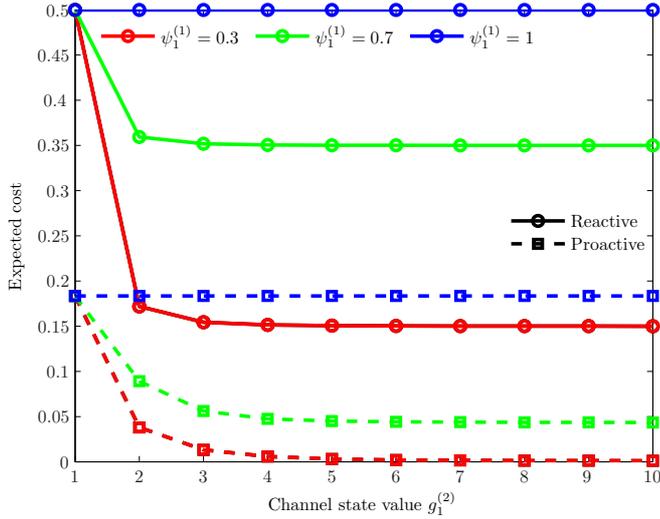


Fig. 2. Comparison of reactive and proactive costs for single user scenario under the time-invariant demand and channel statistics model. The channel probability $\psi_1^{(1)}$ and channel state $g_1^{(2)}$ are varied with fixed user demand probability to $\bar{\pi}_1 = 0.5$ and fixed $g_1^{(1)} = 1$.

For this scenario, the demand probability is set to $\bar{\pi}_1 = 0.5$ and the bad channel state is set to $g_1^{(1)} = 1$. When $\psi_1^{(1)} = 1$ (shown in blue), which means the user always observes $g_1^{(1)}$, there is no impact of $g_1^{(2)}$ on the expected cost for both the schemes. For $\psi_1^{(1)} = 0.3$ and $\psi_1^{(1)} = 0.7$, the cost decreases with increase in $g_1^{(2)}$. This is expected, as when one of the channel states becomes good, the applied proactive service is shifted to that channel condition to minimize the cost. It should be emphasized that highest gains in terms of reduced cost is observed when $g_1^{(2)}$ is twice $g_1^{(1)}$, while beyond this point the reduction in cost is minimal. This effect is attributed to the fourth-order polynomial of the demand cost function. For lower-order polynomial cost functions, the expected cost will reduce gradually for $g_1^{(2)} > 2$.

D. Behavior of the proactive controls

The proactive controls of the single user when there is no demand requests are $\tilde{\mu}_1(0, g_1^{(1)})$ and $\tilde{\mu}_1(0, g_1^{(2)})$ for the two channels states, respectively, and the corresponding controls with demand requests are $\tilde{\mu}_1(1, g_1^{(1)})$ and $\tilde{\mu}_1(1, g_1^{(2)})$. Fig. 3 depicts proactive controls of a single user when the demand probability is $\bar{\pi}_1 = 0.7$ and with equiprobable ($\psi_1^{(1)} = 0.5$) channel states $g_1^{(1)} = 1$, and $g_1^{(2)} = 2$. For this scenario, we note that irrespective of channel state, proactive service is always applied when there is no demand from the user. However, the applied proactive service increases when the channel conditions are better. Proactive service is not applied under bad channel conditions when there are requests from the user. On the other hand, it is advantageous to do proactive service under good channel conditions even when there are demand requests from the user.

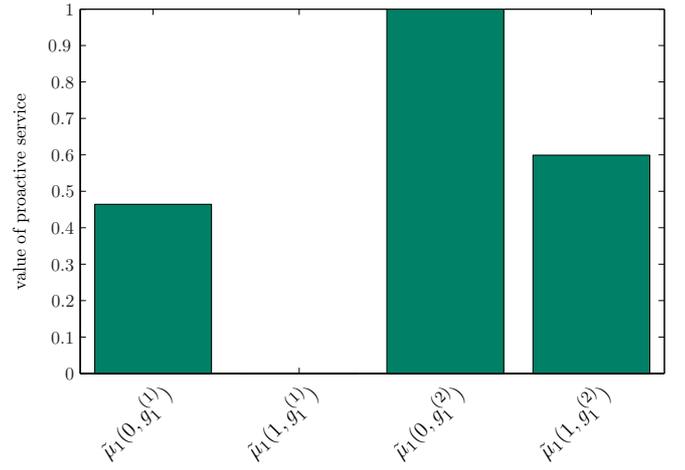


Fig. 3. Proactive controls of the single user with channel states $g_1^{(1)} = 1$ (bad channel state), $g_1^{(2)} = 2$ (good channel state) are (i) $\tilde{\mu}_1(0, g_1^{(1)})$ when no demand request and with channel state $g_1^{(1)}$, (ii) $\tilde{\mu}_1(1, g_1^{(1)})$ with demand request and channel state $g_1^{(1)}$, (iii) $\tilde{\mu}_1(0, g_1^{(2)})$ when no demand request and with channel state $g_1^{(2)}$, and (iv) $\tilde{\mu}_1(1, g_1^{(2)})$ with demand request and channel state $g_1^{(2)}$. Both channel states are equiprobable, i.e., $\psi_1^{(1)} = 0.5$ and user demand probability is set to $\bar{\pi}_1 = 0.7$.

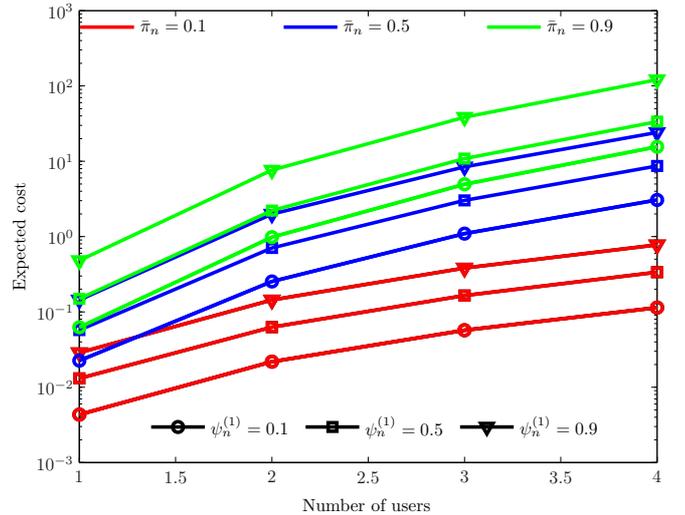


Fig. 4. Impact of number of users on the proactive cost under the time-invariant demand and channel statistics model. The demand probability, channel probability, and the two channel states are kept same for all the users.

E. Impact of number of users on the expected cost

The impact of increasing the number of users in the system on the expected cost is depicted in Fig. 4. For easier analysis of the results, we kept the demand probability, channel probability, and the two channel states to be same for all the users. It can be observed that increase in number of users increases the expected cost, which is due to higher network load with more users. Similar to the single user case (see Fig. 1), the expected cost increases with increase in demand and channel probabilities.

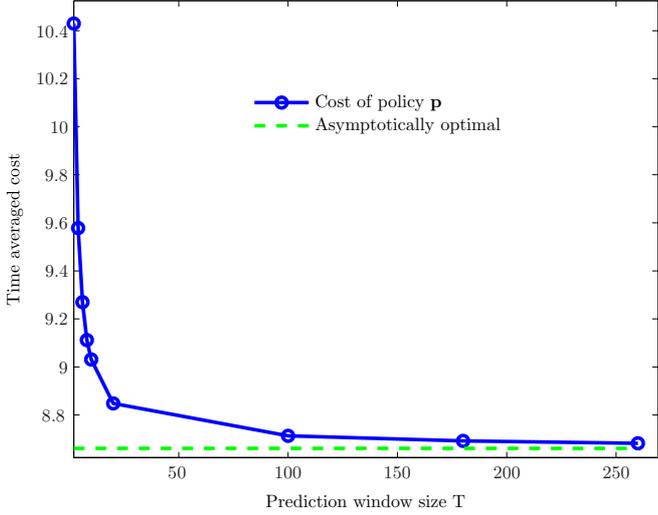


Fig. 5. Impact of proactive window size on achievable cost.

F. Impact of prediction window size on the expected cost

In Fig. 5, we plot the achieved time average cost under \mathbf{p} against the prediction window size T under time-invariant demand and channel statistics. The considered scenario consists of $N = 4$ users in the system who request services based on the demand profile $\bar{\pi}_n = 0.5, n = 1, 2, 3, 4$. We assume equiprobable channel states for all users, i.e., $\psi_n^{(k)} = 0.5, \forall n, k$, with same channel state values $g_n^{(1)} = 1, g_n^{(2)} = 2$, for all the users. It can be observed easily that the policy \mathbf{p} converges rapidly with T to the established lower bound $\underline{c}_{\mathcal{U}}(\bar{\pi}, \Psi)$. For the considered scenario, the policy converges to the lower bound for $T = 250$. We can consider that the service provider can proactively serve requests up to one day ahead. In which the day is divided into T slots, which means $T = 250$ corresponds to a slot size of around 6 minutes which is reasonable for a user to generate one data request.

V. CONCLUSIONS

We studied the impact of demand and channel uncertainties on the design of a proactive scheduler under time-invariant demand and channel statistics models. We have established fundamental lower bound on the achievable cost through proactive scheduling, and developed asymptotically optimal policy that attain the lower bound rapidly as the proactive scheduling window size increases. We observed that the proactive resource scheduler provides better performance in terms of lower achievable cost, compared to reactive scheduler. The proactive scheduler offers better flexibility in scheduling, and adjusts the loads based on user demand requests and channel conditions. We showed that the proactive scheduler depends not only on user demand statistics but also on the realized channel qualities and its associated statistics.

APPENDIX I PROOF OF THEOREM 1

Let us denote $\{\mathbf{u}_{n,t}^*\}_t$ be the optimal proactive scheduling policy of user n under time-invariant data demand and

channel statistics, where $\mathbf{u}_{n,t}^* = (u_{n,t}(\tau))_{\tau}$. The objective of the proactive scheduler is

$$c_T^{\mathcal{P}}(\bar{\pi}, \Psi) = \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \mathbb{E} \left[C_d \left(\sum_{n=1}^N L_{n,l}^{\mathcal{P}}(\mathbf{u}_{n,l}, g_{n,l}) \right) \right]. \quad (11)$$

By joint conditioning on all possible sets of requesting users \mathcal{B}_l and their possible experiencing channel state realizations \mathbf{g}_l at time $l \geq 0$, we can write $c_T^{\mathcal{P}}(\bar{\pi}, \Psi)$ as

$$c_T^{\mathcal{P}}(\bar{\pi}, \Psi) = \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\mathbf{g} \in \mathcal{C}} \sum_{\mathcal{B} \subseteq \mathcal{N}} P(\mathcal{B}_l = \mathcal{B}, \mathbf{g}_l = \mathbf{g}) \times \mathbb{E} \left[C_d \left(\sum_{n=1}^N L_{n,l}^{\mathcal{P}}(\mathbf{u}_{n,l}, g_{n,l}) \right) \middle| \mathcal{B}_l = \mathcal{B}, \mathbf{g}_l = \mathbf{g} \right]. \quad (12)$$

As the set of requesting users \mathcal{B}_l in a time slot l is independent to their observed channel realization \mathbf{g}_l , we can write $P(\mathcal{B}_l = \mathcal{B}, \mathbf{g}_l = \mathbf{g})$ as product of $P(\mathbf{g}_l = \mathbf{g})$ and $P(\mathcal{B}_l = \mathcal{B})$. Then,

$$c_T^{\mathcal{P}}(\bar{\pi}, \Psi) = \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\mathbf{g} \in \mathcal{C}} P(\mathbf{g}_l = \mathbf{g}) \sum_{\mathcal{B} \subseteq \mathcal{N}} P(\mathcal{B}_l = \mathcal{B}) \times \mathbb{E} \left[C_d \left(\sum_{n=1}^N L_{n,l}^{\mathcal{P}}(\mathbf{u}_{n,l}, g_{n,l}) \right) \middle| \mathcal{B}_l = \mathcal{B}, \mathbf{g}_l = \mathbf{g} \right]. \quad (13)$$

Now, incorporating the definition of $L_{n,l}^{\mathcal{P}}(\mathbf{u}_{n,l}, g_{n,l})$, we have

$$c_T^{\mathcal{P}}(\bar{\pi}, \Psi) = \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\mathbf{g} \in \mathcal{C}} P(\mathbf{g}_l = \mathbf{g}) \sum_{\mathcal{B} \subseteq \mathcal{N}} P(\mathcal{B}_l = \mathcal{B}) \times \mathbb{E} \left[C_d \left(\sum_{n=1}^N \left(S C_c(g_{n,l}) - \sum_{\tau=1}^T u_{n,l-\tau}(\tau) C_c(g_{n,l}) \right) + \sum_{\tau=1}^T u_{n,l}(\tau) C_c(g_{n,l}) \right) \middle| \mathcal{B}_l = \mathcal{B}, \mathbf{g}_l = \mathbf{g} \right]. \quad (14)$$

Note that $\{\mathbf{g}_l\}_l$ is an i.i.d. sequence under the time-invariant channel statistics model, thus we could use $P_c(\mathbf{g}) = \prod_n \sum_{k=1}^{K_n} \psi_n^{(k)} \mathbb{I}(g_n = g_n^{(k)})$, where $\mathbb{I}(g_n = g_n^{(k)})$ is an indicator function that equals one if and only if $g_n = g_n^{(k)}$. That means the set of channel states of users in one time slot are independent in another time slot. Furthermore, $\{\mathcal{B}_l\}_l$ is an i.i.d. sequence under the time-invariant demand statistics model, thus we could use $P_d(\mathcal{B}) = \prod_{n \in \mathcal{B}} \bar{\pi}_n \prod_{m \notin \mathcal{B}} (1 - \bar{\pi}_m)$. So, the set of user requests in one time slot are independent in another time slot. We obtain

$$c_T^{\mathcal{P}}(\bar{\pi}, \Psi) = \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\mathbf{g} \in \mathcal{C}} P_c(\mathbf{g}) \sum_{\mathcal{B} \subseteq \mathcal{N}} P_d(\mathcal{B}) \times \mathbb{E} \left[C_d \left(\sum_{n=1}^N \left(S C_c(g_n) - \sum_{\tau=1}^T u_{n,l-\tau}(\tau) C_c(g_n) \right) + \sum_{\tau=1}^T u_{n,l}(\tau) C_c(g_n) \right) \middle| \mathcal{B}, \mathbf{g} \right]. \quad (15)$$

We can apply Jensen's inequality on (15), since $C_d(x)$ is assumed to be strictly convex. Moreover, \mathcal{B}_l is independent of $\sum_{\tau=1}^T u_{n,l-\tau}(\tau)$, because the current set of users \mathcal{B}_l does not influence the past services in the time slots $\{l-1, l-2, \dots, l-T\}$. However, they will influence the future services because the future services are dependent on the load in the current time slot. Similarly, \mathbf{g}_l is independent of $\sum_{\tau=1}^T u_{n,l-\tau}(\tau)$. The current channel state of the users does not influence past applied proactive controls but on the future proactive services. We can write

$$\begin{aligned} c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) &\geq \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\mathbf{g} \in \mathcal{C}} P_c(\mathbf{g}) \sum_{\mathcal{B} \subseteq \mathcal{N}} P_d(\mathcal{B}) \\ &\quad \times C_d \left(\sum_{n \in \mathcal{B}} \left(S C_c(g_n) - \sum_{\tau=1}^T \mathbb{E}[u_{n,l-\tau}^*(\tau)] C_c(g_n) \right) \right. \\ &\quad \left. + \sum_{n=1}^N \sum_{\tau=1}^T \mathbb{E}[u_{n,l}^*(\tau) | \mathcal{B}, \mathbf{g}] C_c(g_n) \right). \end{aligned} \quad (16)$$

Since $\sum_{\mathcal{B} \subseteq \mathcal{N}} P_d(\mathcal{B}) = 1$, $\sum_{\mathbf{g} \in \mathcal{C}} P_c(\mathbf{g}) = 1$, and $\frac{1}{t} \sum_{l=0}^{t-1} 1 = 1$, we can apply Jensen's inequality again. Also invoking $\limsup_{t \rightarrow \infty} (-f(t)) = -\liminf_{t \rightarrow \infty} (f(t))$, we have

$$\begin{aligned} c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) &\geq \sum_{\mathbf{g} \in \mathcal{C}} P_c(\mathbf{g}) \sum_{\mathcal{B} \subseteq \mathcal{N}} P_d(\mathcal{B}) C_d \left(\sum_{n \in \mathcal{B}} \left(S C_c(g_n) \right. \right. \\ &\quad \left. \left. - \liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\tau=1}^T \mathbb{E}[u_{n,l-\tau}^*(\tau)] C_c(g_n) \right) \right. \\ &\quad \left. + \sum_{n=1}^N C_c(g_n) \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\tau=1}^T \mathbb{E}[u_{n,l}^*(\tau) | \mathcal{B}, \mathbf{g}] \right). \end{aligned} \quad (17)$$

As $C_d(x)$ is monotonically increasing in x , replacing \limsup on the right hand side of the expression by \liminf , which yields

$$\begin{aligned} c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) &= \sum_{\mathbf{g} \in \mathcal{C}} P_c(\mathbf{g}) \sum_{\mathcal{B} \subseteq \mathcal{N}} P_d(\mathcal{B}) C_d \left(\sum_{n \in \mathcal{B}} \left(S C_c(g_n) \right. \right. \\ &\quad \left. \left. - \liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\tau=1}^T \mathbb{E}[u_{n,l-\tau}^*(\tau)] C_c(g_n) \right) \right. \\ &\quad \left. + \sum_{n=1}^N C_c(g_n) \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\tau=1}^T \mathbb{E}[u_{n,l}^*(\tau) | \mathcal{B}, \mathbf{g}] \right). \end{aligned} \quad (18)$$

We define $\tilde{\mu}_n(\mathcal{B}, \mathbf{g}) = \liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\tau=1}^T \mathbb{E}[u_{n,l}^*(\tau) | \mathcal{B}, \mathbf{g}]$. Then, we can rewrite $\liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\tau=1}^T \mathbb{E}[u_{n,l-\tau}^*(\tau)]$ in the following

way

$$\begin{aligned} &\liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\tau=1}^T \mathbb{E}[u_{n,l-\tau}^*(\tau)] \\ &= \liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\tau=1}^T \sum_{\mathbf{h} \in \mathcal{C}} \sum_{\mathcal{D} \subseteq \mathcal{N}} P(\mathbf{h}_l = \mathbf{h}) P(\mathcal{D}_l = \mathcal{D}) \\ &\quad \times \mathbb{E}[u_{n,l-\tau}^*(\tau) | \mathbf{h}_l = \mathbf{h}, \mathcal{D}_l = \mathcal{D}] \\ &= \sum_{\mathbf{h} \in \mathcal{C}} P_c(\mathbf{h}) \sum_{\mathcal{D} \subseteq \mathcal{N}} P_d(\mathcal{D}) \left(\liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \sum_{\tau=1}^T \mathbb{E}[u_{n,l-\tau}^*(\tau) | \mathbf{h}, \mathcal{D}] \right) \\ &= \sum_{\mathbf{h} \in \mathcal{C}} P_c(\mathbf{h}) \sum_{\mathcal{D} \subseteq \mathcal{N}} P_d(\mathcal{D}) \tilde{\mu}_n(\mathcal{D}, \mathbf{h}) \end{aligned}$$

and we obtain

$$\begin{aligned} c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) &\geq \sum_{\mathbf{g} \in \mathcal{C}} P_c(\mathbf{g}) \sum_{\mathcal{B} \subseteq \mathcal{N}} P_d(\mathcal{B}) C_d \left(\sum_{n \in \mathcal{B}} \left(S C_c(g_n) \right. \right. \\ &\quad \left. \left. - \left(\sum_{\mathbf{h} \in \mathcal{C}} P_c(\mathbf{h}) \sum_{\mathcal{D} \subseteq \mathcal{N}} P_d(\mathcal{D}) \tilde{\mu}_n(\mathcal{D}, \mathbf{h}) \right) C_c(g_n) \right) \right. \\ &\quad \left. + \sum_{n=1}^N \tilde{\mu}_n(\mathcal{B}, \mathbf{g}) C_c(g_n) \right). \end{aligned} \quad (19)$$

It should be noted that constraints on the proactive service (3) and (4) implies $0 \leq \tilde{\mu}_n(\mathcal{B}, \mathbf{g}) \leq S$, $\forall n, \mathcal{B}, \mathbf{g}$. Now, by minimizing the right-hand-side of the last expression over all feasible choices of $\{\tilde{\mu}_n(\mathcal{B}, \mathbf{g})\}_{n, \mathcal{B}, \mathbf{g}}$, the theorem is proved.

APPENDIX II PROOF OF THEOREM 2

It suffices to prove that $\limsup_{T \rightarrow \infty} c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) = \liminf_{T \rightarrow \infty} c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi})$. We start by $\limsup_{T \rightarrow \infty} c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi})$. Since policy \mathbf{p} is stationary, we can write $c_T^{\mathcal{P}}(\bar{\boldsymbol{\pi}}, \boldsymbol{\Psi}) =$

$$\begin{aligned} &\mathbb{E} \left[C_d \left(\sum_{n=1}^N \left(S - \sum_{\tau=1}^T u_{n,t-\tau}(\tau) \right) d_{n,t} C_c(g_{n,t}) \right. \right. \\ &\quad \left. \left. + \sum_{n=1}^N \sum_{\tau=1}^T u_{n,t}(\tau) C_c(g_{n,t}) \right) \right] = \sum_{\mathcal{B} \subseteq \mathcal{N}} P_d(\mathcal{B}) \sum_{\mathbf{g} \in \mathcal{C}} P_c(\mathbf{g}) \times \\ &\quad \mathbb{E} \left[C_d \left(\sum_{n \in \mathcal{B}} \left(S - \sum_{\tau=1}^T u_{n,t-\tau}(\tau) \right) C_c(g_n) \right. \right. \\ &\quad \left. \left. + \sum_{n=1}^N \mu_n(\mathcal{B}, \mathbf{g}) \right) | \mathcal{B}_t = \mathcal{B}, \mathbf{g}_t = \mathbf{g} \right]. \end{aligned}$$

Now, we consider the sum $\sum_{\tau=1}^T u_{n,t-\tau}(\tau)$, which is independent of $\mathcal{B}_t, \mathbf{g}_t$. Define a random variable $Z_T(\mathcal{D}, \mathbf{h})$ which counts the number of occurrences of the pair of requesting set $\mathcal{D} \subseteq \mathcal{N}$ and associated channel gain vector $\mathbf{h} \in \mathcal{C}$, in slots $t-T, \dots, t-1$. Then, $\sum_{\tau=1}^T u_{n,t-\tau}(\tau) = \sum_{\mathcal{D} \subseteq \mathcal{N}} \sum_{\mathbf{h} \in \mathcal{C}} \frac{\mu_n(\mathcal{D}, \mathbf{h}) Z_T(\mathcal{D}, \mathbf{h})}{T}$. By the strong law of large

numbers,

$$\limsup_{T \rightarrow \infty} \sum_{\mathcal{D} \subseteq \mathcal{N}} \sum_{\mathbf{h} \in \mathcal{C}} \frac{\mu_n(\mathcal{D}, \mathbf{h}) Z_T(\mathcal{D}, \mathbf{h})}{T} = \mu_n(\mathcal{D}, \mathbf{h}) P_d(\mathcal{D}) P_c(\mathbf{h}), \quad \text{w.p. 1.}$$

By noting that the system load at any time slot is uniformly bounded above, bounded convergence theorem implies

$$\begin{aligned} \limsup_{T \rightarrow \infty} c_T^{\mathbf{p}}(\bar{\pi}, \Psi) &= \sum_{\mathcal{B} \subseteq \mathcal{N}} P_d(\mathcal{B}) \sum_{\mathbf{g} \in \mathcal{C}} P_c(\mathbf{g}) C_d \left(\sum_{n \in \mathcal{B}} (S - \right. \\ &\quad \left. \sum_{\mathcal{D} \subseteq \mathcal{N}} \sum_{\mathbf{h} \in \mathcal{C}} \mu_n(\mathcal{D}, \mathbf{h}) P_d(\mathcal{D}) P_c(\mathbf{h})) C_c(g_n) \right. \\ &\quad \left. + \sum_{n=1}^N \mu_n(\mathcal{B}, \mathbf{g}) C_c(g_n) \right) = \underline{c}_l(\bar{\pi}, \Psi). \end{aligned}$$

Thus we have established that average expected cost under policy \mathbf{p} attains the global lower bound as prediction window size grows to infinity. Now by the definition of $c_T^{\mathbf{p}}(\bar{\pi}, \Psi)$ being the minimum possible cost achieved by proactive scheduling with prediction window T , it follows that $\limsup_{T \rightarrow \infty} c_T^{\mathbf{p}}(\bar{\pi}, \Psi) = \liminf_{T \rightarrow \infty} c_T^{\mathbf{p}}(\bar{\pi}, \Psi)$.

REFERENCES

- [1] "Ericsson mobility report," <http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf>, June 2015.
- [2] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [3] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "Human mobility, social ties, and link prediction," in *17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, pp. 1100–1108.
- [4] B. S. Jensen, J. E. Larsen, K. Jensen, J. Larsen, and L. K. Hansen, "Estimating human predictability from mobile sensor data," in *IEEE International Workshop on Machine Learning for Signal Processing*, 2010, pp. 196–201.
- [5] R. Di Taranto, L. S. Muppisetty, R. Raulefs, D. Stock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 102–112, Nov 2014.
- [6] L. S. Muppisetty, T. Svensson, and H. Wymeersch, "Spatial wireless channel prediction under location uncertainty," *Accepted in IEEE Transactions on Wireless Communications*, 2015.
- [7] M. Malmirchegini and Y. Mostofi, "On the spatial predictability of communication channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 964–978, 2012.
- [8] S.-J. Kim, E. Dall'Anese, and G. Giannakis, "Cooperative spectrum sensing for cognitive radios using kriged Kalman filtering," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 24–36, 2011.
- [9] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.
- [10] J. Tadrous and A. Eryilmaz, "On optimal proactive caching for mobile networks with demand uncertainties," *Accepted in IEEE/ACM Transactions on Networking*, 2015.
- [11] J. Tadrous, A. Eryilmaz, and H. El Gamal, "Proactive resource allocation: Harnessing the diversity and multicast gains," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 4833–4854, 2013.
- [12] L. Huang, S. Zhang, M. Chen, and X. Liu, "When backpressure meets predictive scheduling," in *Proceedings of the 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2014, pp. 33–42.
- [13] S. Zhang, L. Huang, M. Chen, and X. Liu, "Effect of proactive serving on user delay reduction in service systems," in *The 2014 ACM International Conference on Measurement and Modeling of Computer Systems*, 2014, pp. 573–574.
- [14] J. Lee and N. Jindal, "Asymptotically optimal policies for hard-deadline scheduling over fading channels," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2482–2500, 2013.
- [15] M. Zafer and E. Modiano, "Minimum energy transmission over a wireless channel with deadline and power constraints," *IEEE Transactions on Automatic Control*, vol. 54, no. 12, pp. 2841–2852, Dec 2009.