

Proactive Caching for Energy-Efficiency in Wireless Networks: A Markov Decision Process Approach

Zhijie Chen

Department of Electrical Engineering
Stanford University
Stanford, CA, USA, 94305
Email: zcchen@stanford.edu

Hoshyar Mohammed, and Wei Chen, *Senior Member, IEEE*

Department of Electronic Engineering / TNLlist
Tsinghua University
Beijing, China, 100084

Email: mu-h16@mails.tsinghua.edu.cn, wchen@tsinghua.edu.cn

Abstract—Content caching in wireless networks provides a substantial opportunity to trade off low cost memory storage with energy consumption, yet finding the optimal causal policy with low computational complexity remains a challenge. This paper models the Joint Pushing and Caching (JPC) problem as a Markov Decision Process (MDP) and provides a solution to determine the optimal randomized policy. A novel approach to decouple the influence from buffer occupancy and user requests is proposed to turn the high-dimensional optimization problem into three low-dimensional ones. Furthermore, a non-iterative algorithm to solve one of the sub-problems is presented, exploiting a structural property we found as *generalized monotonicity*, and hence significantly reduces the computational complexity. The result attains close performance in comparison with theoretical bounds from non-practical policies, while benefiting from higher time efficiency than the unadapted MDP solution.

I. INTRODUCTION

With the escalating growth of mobile data traffic caused by the proliferation of smart mobile devices, and the growing online services (i.e., Video on Demand (VoD) streaming, Facebook, Twitter,...), corresponding energy consumption is increasing considerably [1]. Furthermore, video streaming and social applications require high bandwidth and strict delay constraint, which in turn has further adverse effects on user experience quality and Operational Expenses (OpEx). In addition, It has been reported that the network data traffic will continue growing in the future years [1]. Hence, efficient energy utilization is of paramount importance in the design of future wireless networks. In this paper, our objective is to exploit the cache-enabled user devices to reduce the energy consumption by focusing particularly on the wireless transmission cost.

The massive mobile users generate issues to be addressed, such as, extremely high throughput and stringent Quality of Service (QoS) requirements, and excessive energy consumption. One way to deal with such issues is to deploy larger number of small cells and fog style access points [2], [3]. However, this approach does not address the energy efficiency and still large undesirable latency and network congestion could be induced during peak-traffic hours. Hence, content caching at the edge of wireless networks has attracted much attention from both academia and industry [4]–[8].

This research was supported by the National Science Foundation of China under Grant Nos. 61671269 and 61322111, and the National 973 Program of China under Project No.2013CB336600.

Moving contents to the edge of network emerged as a prospective technique, that can significantly reduce transmission latency [4], [5], control traffic load [6], and reduce energy consumption [7], or can be employed to increase overall system throughput [8]. Content caching utilizes the recent advances in the field of context awareness and leverages on the low price memory storage to enhance the system performance. Content caching enables proactive transmission (i.e., *pre-serve*) that allows the user to download content files over a longer period, hence, reducing energy consumption. It can also mitigate severe network condition (i.e., peak-traffic hours) by pushing at favorable transmission times. For instance, by exploiting the user demand information the base station (BS) can push the desired content files prior to the playback time (i.e., in video streaming). Hence, saving a significant amount of energy by avoiding the unfavorable channel conditions. It also provides another advantage by shifting traffic load (or equivalently reducing traffic variability) for better system utilization. In another line of research, the objective is to increase the throughput of content-centric wireless networks to better support the exponential growth of wireless data traffic. In [8] the available cache memory at the user is exploited via a joint pushing and caching method to increase the system throughput under non-causal, statistical, and causal knowledge of user request delay information (RDI).

In this paper, we consider an optimization problem of a proactive caching wireless communications channel, with limited data buffer capacity available at the receiver end. Specifically, we formulate the joint pushing and caching (JPC) problem as an infinite horizon average cost Markov decision process (MDP) and devise a randomized policy to minimize the average energy consumption over time. In each timeslot the amount of data to transmit is a random variable, whose distribution is decided by the optimal policy, taking both the buffer occupancy and user requests into account. Numerically solving the optimization problem is computationally demanding due to the two-dimensional state space and randomized policy, known as the curse of dimensionality. Therefore, we develop a novel approach to decouple the factors of buffer occupancy and user requests by introducing the degenerated state space, and hence breakdown the optimization into three more tractable sub-problems. Furthermore, we found a structural

property that gives a non-iterative way to design the optimal policy under certain constraints in the degenerated space. We name the property as *generalized monotonicity*, which brings significant improvement to the computational complexity. The model in this work performs well in comparison against two theoretical bounds, which were derived from assuming non-causal knowledge of user request delay information (RDI) [9] and unlimited cache capacity, respectively.

II. SYSTEM MODEL

Consider a wireless communication system between a server and a user equipped with a limited buffer as depicted in Fig. 1. The system operates over an infinite time horizon in a discrete time fashion, with timeslots $t = 0, 1, 2, \dots$. At the beginning of each timeslot, the user requests for a certain amount of data, which must be fulfilled by the end of the timeslot. The data may be transmitted in advance, stored and read from the buffer (proactively), or be transmitted on-demand (reactively), or combined. Also, we consider the scenario where data stream has been temporally ordered, i.e., the server knows what to be requested in the near future, while not knowing how soon the requests will be made.

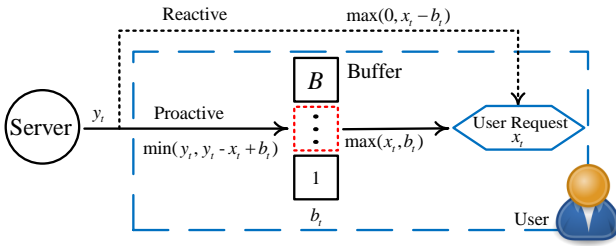


Fig. 1. System model. As explained in section (II-A), fetching data from the buffer always has higher priority over via on-demand transmission, based on the assumption of temporal order. This observation gives the data flow on each edge.

A. Request model and buffer state

We assume that data is requested in content items of identical size. Thus, we may denote the number of content items requested in timeslot t as $x_t \in \mathcal{X} \triangleq \{0, 1, \dots, X\}$, where X is the maximum possible value of x_t . We further assume that $x_t, t = 0, 1, \dots$ are *i.i.d.* integer random variables bounded by X , yielding probability mass function (p.m.f) $f_X(\cdot)$. One can always limit x_t to integer values by choosing a small enough size for content items. Let a $(X+1)$ -dimensional vector \mathbf{p} denotes the p.m.f. of x_t , i.e., $p_x = f_X(x), x \in \mathcal{X}$.

Let buffer state b_t denote the number of content items at the beginning of timeslot t , $b_t \in \mathcal{B} \triangleq \{0, 1, \dots, B\}$. This implies the buffer storage capacity being B content items. Since data stream is well-ordered, for any two content items, we always know which one would be requested first, and thus which to proactively push first. Therefore, as long long as the buffer is not empty, there is no reason to transmit data on-demand. In other words, on-demand transmission happens if and only if $x_t > b_t$.

B. System State Model

The system state space is denoted by \mathcal{S} which consists of b and x elements and has the cardinality of $(B+1) \times (X+1)$. The pair (b_t, x_t) constitutes each state $s_t \in \mathcal{S}$. Furthermore, we define degenerated state $\mathbf{b}_b = \{s : s = (b, x), x \in \mathcal{X}\}$, and degenerated state space $\mathcal{B} = \{\mathbf{b}_b : b \in \mathcal{B}\}$. When it does not cause confusion, we use \mathbf{b}_b and b interchangeably.

Let y_t denote the number of content items downloaded within timeslot t . Under a given policy, y_t is a random variable with distribution being contingent on buffer occupancy b_t and user requests x_t , i.e., s_t . Buffer occupancy evolves as a time-homogeneous Markov chain which can be described as

$$b_{t+1} = b_t + y_t - x_t, \quad (1)$$

where

$$0 \leq b_t \leq B, \quad \forall t. \quad (2)$$

the first inequality in (2) comes from the model setting that requests must be fulfilled within the requested timeslot t , and thus $b_t + y_t \geq x_t$. The Fig. 2 depicts the dynamical relation between y_t , b_t and x_t in each timeslot. A policy π would determine the distribution of y_t from b_t and x_t .

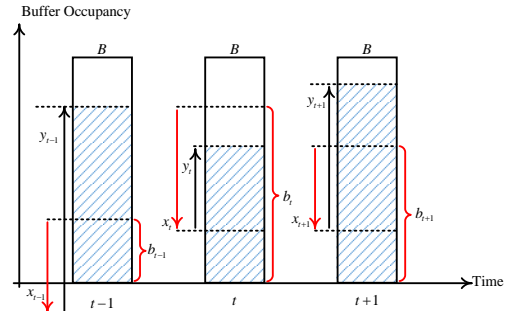


Fig. 2. Buffer state evolution. An illustration of the relationship between the buffer occupancy b_t , user requests x_t , and data transmission y_t in a timeslot.

We consider a causal knowledge of user RDI, that the user requests for a future timeslot stays unknown until the beginning of the timeslot.

The transmission action in each timeslot incurs a certain energy cost, whose corresponding power is typically a convex and continuous function of data rate. Conventionally we assume this function to be exponential [10]. By [9], constant transmission rate should be used within a timeslot to minimize energy consumption. Absorbing all constants and normalizing the scalar in our model, we give the energy consumption in timeslot t by:

$$\rho_t = \rho(y_t) = \eta^{y_t} - 1, \quad (3)$$

where ρ_t is the energy consumption and $\eta > 1$ is a constant. Since, y_t takes finitely many values, it can be shown that $\mathbb{E}\rho_t < \infty$.

III. A RANDOMIZED MDP PROBLEM FORMULATION

In a randomized policy, a transmission action y_t yields a distribution which we design contingent on s_t , i.e., $y_t \sim f_{Y|s_t}(y)$. The optimal transmission policy selects a distribution for y_t .

However, the range where random variable y_t takes non-zero probability changes given different s_t . By (1), $\sigma(y_t | s_t) = \sigma(b_{t+1} | s_t)$. Therefore, we study the conditional probability of b_{t+1} instead of y_t . The system evolves as depicted in Fig. 3.

Let a $(B+1)$ -dimensional vector $\mathbf{d}^{s,\pi}$ denote a randomized decision, i.e., p.m.f., conditioning on state s under policy π , where $d_{b^+}^{s,\pi} \triangleq \Pr(b_{t+1} = b^+ | s_t = s, \pi)$. All $(\mathbf{d}^{s,\pi})_{s \in \mathcal{S}}$ give the transition probabilities in \mathcal{S} and \mathfrak{B} . Denote the transition probability matrix in \mathfrak{B} under policy π by $\mathbf{A}^\pi \triangleq [\mathbf{a}^{0,\pi}, \mathbf{a}^{1,\pi}, \dots, \mathbf{a}^{B,\pi}]^\top$, where $\mathbf{a}^{b,\pi} = \sum_x \mathbf{d}^{(b,x),\pi} p_x$. Since

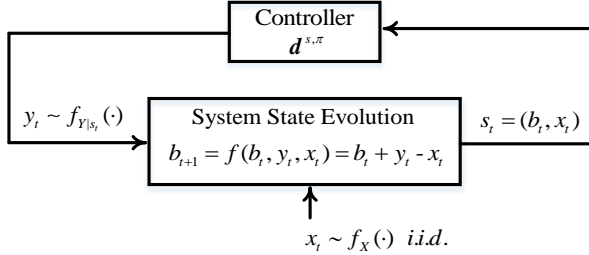


Fig. 3. At each timeslot t the controller, i.e., the server, observes the current system state s_t and applies a control $y_t \sim f_{Y|S_t}(\cdot)$ (and equivalently, $b_{t+1} \sim \mathbf{d}^{s_t,\pi}$) contingent on the state.

every state in \mathcal{S} has to be assigned with a randomized decision, the corresponding policy space is $(B+1)^2 \times (X+1)$ -dimensional. Denote the probability that the system occupies state $(b, x) \in \mathcal{S}$ in timeslot t under policy π as $q_{b,x}^{t,\pi} \triangleq \Pr(s_t = (b, x) | \pi)$, and define matrix $\mathbf{Q}^{t,\pi} \triangleq (q_{b,x}^{t,\pi})_{b \in \mathcal{B}, x \in \mathcal{X}}$, which contains the probabilities for all states. Similarly, we denote the state probability of $b_b \in \mathfrak{B}$ in timeslot t under policy π as $r_b^{t,\pi} \triangleq \Pr\{b_t = b | \pi\}$, and define vector $\mathbf{r}^{t,\pi} \triangleq (r_b^{t,\pi})_{b \in \mathcal{B}}$. Since, $x_t \sim f_X(\cdot)$ is i.i.d., we have

$$\mathbf{Q}^{t,\pi} = \mathbf{r}^{t,\pi} \mathbf{p}^\top, \quad \forall t > 0, \pi. \quad (4)$$

Under a given policy π , the expected energy cost in timeslot t solely depends on state s_t . We define the expected energy cost for state $s = (b, x)$ as $\omega_b^\pi \triangleq \mathbb{E}[\rho_t | s_t = s, \pi]$. Let $\Phi^B \triangleq [1, \eta, \eta^2, \dots, \eta^B]^\top$ and $\Phi^X \triangleq [1, \eta, \eta^2, \dots, \eta^X]^\top$, we have

$$\omega_s^\pi = \eta^{x-b} \sum_{m=0}^B \eta^m d_m^{s,\pi} - 1 = \eta^{x-b} \Phi^{B\top} \mathbf{d}^{s,\pi} - 1. \quad (5)$$

denote with matrix $\Omega^\pi \triangleq (\omega_{b,x}^\pi)_{b \in \mathcal{B}, x \in \mathcal{X}}$ the expected energy costs for all states. Under the average reward criterion, we formally state the optimization problem as

$$\underset{\pi}{\text{minimize}} \quad \mathcal{L}(\pi) = \mathbf{r}^{\infty,\pi\top} \Omega^\pi \mathbf{p} \quad (6a)$$

$$\text{subject to} \quad \mathbf{A}^{\pi\top} \mathbf{r}^{\infty,\pi} = \mathbf{r}^{\infty,\pi}, \quad (6b)$$

$$\mathbf{d}^{s,\pi} \geq 0, \quad \forall s, \quad (6c)$$

$$\mathbf{1}^\top \mathbf{d}^{s,\pi} = 1, \quad \forall s, \quad (6d)$$

$$d_m^{(b,x),\pi} = 0, \quad \forall m < b - x. \quad (6e)$$

where constraint (6b) implies stationary distribution and (6c) and (6d) ensure that $\mathbf{d}^{s,\pi}$ corresponds to a p.m.f.. The last

constraint rules out the possibility of y_t being negative, since the number of content item transmitted must be positive or 0.

As the policy space is $(B+1)^2 \times (X+1)$ -dimensional, problem (6) is an optimization with high dimensionality, whose convexity is hard to determine. Value iteration being a popular algorithm regarding solving MDP problems, overwhelming computational complexity is yet a problem if to run value iteration in \mathcal{S} . We found a novel approach to decouple the influence from b_t and x_t and hence turn (6) to three sub-problems, which are low-dimensional and more tractable. By solving the three sub-problems specified below, we equivalently solves (6).

First, we assume a transition matrix in \mathfrak{B} is given as \mathbf{A} and find the optimal policy, i.e., $\{\mathbf{d}^{s,\pi} : s \in \mathcal{S}\}$, among all policies that results in \mathbf{A} . Secondly, we find the optimal \mathbf{A} whose corresponding optimal policy minimizes the expected energy consumption in one iteration step. Thirdly, we carry out value iteration in \mathfrak{B} , each step searching for the optimal \mathbf{A} iteratively, until the transition matrix converges to \mathbf{A}^* . By [11], \mathbf{A}^* is the minimizer for average energy cost among all possible \mathbf{A} . We argue that we have also found the optimal π^* , among all policy space, that minimize the average energy cost as

$$\min_{\mathbf{A}} \left(\min_{\pi} \mathcal{L}(\pi) \mid \mathbf{A} \right) = \min_{\pi} \mathcal{L}(\pi). \quad (7)$$

We next elaborate each of the three sub-problems.

A. Optimal Decisions with Given Transition Matrix in \mathfrak{B}

Denote with ω_b^π the expected energy cost under policy π in a timeslot belonging to state b_b . We have

$$\omega_b^\pi \triangleq \mathbb{E}[\rho_t | b_t = b, \pi] = \sum_{x=0}^X p_x \omega_{(b,x)}^\pi = \Omega_{(b,:)}^\pi \mathbf{p}, \quad (8)$$

where $\Omega_{(b,:)}^\pi$ denotes the row vector from the b^{th} row of Ω . Recall that $\mathbf{A}^\pi = [\mathbf{a}^{0,\pi} \ \mathbf{a}^{1,\pi} \ \dots \ \mathbf{a}^{B,\pi}]^\top$ and $\mathbf{a}^{b,\pi} = \sum_x \mathbf{d}^{(b,x),\pi}$, i.e., the b^{th} row of \mathbf{A} is solely determined by decisions from b_b . We formulate a matrix that represents decisions from b_b as $\mathbf{D}^{b,\pi} \triangleq [\mathbf{d}^{(b,0),\pi} \ \mathbf{d}^{(b,1),\pi} \ \dots \ \mathbf{d}^{(b,X),\pi}]^\top$. The first sub-problem, finding the optimal decisions under the constraint of a fixed \mathbf{A} , is formally stated as:

$$\underset{\mathbf{D}^{b,\pi}}{\text{minimize}} \quad \omega_b^\pi = \eta^{-b} \Phi^{X\top} \text{diag}(\mathbf{p}) \mathbf{D}^{b,\pi} \Phi^B - 1 \quad (9.a)$$

$$\text{subject to} \quad \mathbf{D}^{b,\pi\top} \mathbf{p} = \mathbf{a}_b^\pi, \quad (9.b)$$

$$\mathbf{D}^{b,\pi} \mathbf{1} = \mathbf{1}, \quad (9.c)$$

$$D_{m,n}^{b,\pi} \geq 0, \quad \forall m, n \quad (9.d)$$

$$D_{m,n}^{b,\pi} = 0, \quad \forall m + n < b, \quad (9.e)$$

for every $b \in \mathcal{B}$. We define a function $h : \mathbb{R}^{X+1} \rightarrow \mathbb{R}$ as:

$$h(\mathbf{a}^{b,\pi}) = \min_{\pi} (\omega_b^\pi | \mathbf{a}^{b,\pi}), \quad (10)$$

in light of the optimization problem (9). In section IV, we propose an efficient algorithm to find $h(\mathbf{a}_b^\pi)$ without iteration, exploiting a certain structural property we name as *generalized monotonicity*.

B. Optimal Transition Probability Matrix in \mathfrak{B}

In a finite MDP with N timeslots, define v_b^t as the expected total cost over period $\{N-t+1, N-t+2, \dots, N\}$, starting from $b_{N-t+1} = b$. Denote vector $\mathbf{v}^t \triangleq (v_b^t)_{b \in \mathcal{B}}$. The iteration from $t-1$ to t can be written as

$$\underset{\mathbf{a}^{b,\pi}}{\text{minimize}} \quad v_b^t = h(\mathbf{a}^{b,\pi}) + \mathbf{a}^{b,\pi^\top} \mathbf{v}^{t-1} \quad (11.a)$$

$$\text{subject to} \quad \mathbf{a}^{b,\pi} \geq 0, \quad (11.b)$$

$$\mathbf{1}^\top \mathbf{a}^{b,\pi} = 1, \quad (11.c)$$

$$a_m^{b,\pi} \leq \sum_{x=b-m}^X p_x, \quad \forall m = 0, 1, \dots, b. \quad (11.d)$$

(11.b) and (11.c) are the nature of p.m.f., and (11.d) is derived from (9.b) and (9.e). Now we prove that optimization problem (11) is convex, and hence can be easily solved by conventional optimization tools, say, the interior-point method.

Proof: Let \mathbf{a}^{b,π_1} and \mathbf{a}^{b,π_2} be two feasible points of (11), whose corresponding solutions in (9) are \mathbf{D}^{b,π_1} and \mathbf{D}^{b,π_2} , respectively. For all $\lambda \in \{\lambda \in \mathbb{R} | 0 < \lambda < 1\}$, let

$$\mathbf{a}^{b,\pi'} = (1-\lambda)\mathbf{a}^{b,\pi_1} + \lambda\mathbf{a}^{b,\pi_2}, \quad (12)$$

and, let

$$\mathbf{D}^{b,\pi'} = (1-\lambda)\mathbf{D}^{b,\pi_1} + \lambda\mathbf{D}^{b,\pi_2}. \quad (13)$$

it can be directly shown that $\mathbf{a}^{b,\pi'}$ and $\mathbf{D}^{b,\pi'}$ satisfy constraint sets (11.b)-(11.d) and (9.c)-(9.e) respectively. Also we have

$$\begin{aligned} \mathbf{D}^{b,\pi'} \mathbf{p} &= ((1-\lambda)\mathbf{D}^{b,\pi_1} + \lambda\mathbf{D}^{b,\pi_2})^\top \mathbf{p} \\ &= (1-\lambda)\mathbf{a}^{b,\pi_1} + \lambda\mathbf{a}^{b,\pi_2} \\ &= \mathbf{a}^{b,\pi'}, \end{aligned} \quad (14)$$

i.e., $\mathbf{D}^{b,\pi'}$ satisfies (9.b) with $\mathbf{a}^{b,\pi'}$. Thus, $\mathbf{a}^{b,\pi'}$ and $\mathbf{D}^{b,\pi'}$ are feasible points of (11) and (9) respectively. Now denote with $\omega_b^{\pi'}$ the expected average energy cost carried out by $\mathbf{D}^{b,\pi'}$. Because (9.a) is linear on $\mathbf{D}^{b,\pi}$, $\omega_b^{\pi'} = (1-\lambda)h(\mathbf{a}^{b,\pi_1}) + \lambda h(\mathbf{a}^{b,\pi_2})$, which gives a feasible value of (9) under $\mathbf{a}^{b,\pi'}$. By definition,

$$\begin{aligned} h(\mathbf{a}^{b,\pi'}) &= \min_{\pi} \left(\omega_b^{\pi} \mid \mathbf{a}^{b,\pi'} \right) \\ &\leq \omega_b^{\pi'} \\ &= (1-\lambda)h(\mathbf{a}^{b,\pi_1}) + \lambda h(\mathbf{a}^{b,\pi_2}). \end{aligned} \quad (15)$$

thus, h is convex. Since that $\mathbf{a}^{b,\pi^\top} \mathbf{v}^{t-1}$ is linear, optimization problem (11) is convex. ■

C. Value Iteration in degenerated state space \mathfrak{B}

By III-A and III-B, we find the optimal transition matrix in \mathfrak{B} and its corresponding policy, in light of one-step iteration. Now we start from:

$$\mathbf{v}^0 = \mathbf{0}, \quad (16)$$

and carry out the optimality equations given by

$$\begin{aligned} v_b^t &= \min_{\mathbf{a}^{b,\pi}} h(\mathbf{a}^{b,\pi}) + \sum_{b^+=0}^B a_{b^+}^{b,\pi} v_{b^+}^{t-1} \\ &= \min_{\mathbf{a}^{b,\pi}} h(\mathbf{a}^{b,\pi}) + \mathbf{a}^{b,\pi^\top} \mathbf{v}^{t-1}, \end{aligned} \quad (17)$$

for $\forall b \in \mathcal{B}$ and $t = 1, 2, \dots$. The transition matrix A^π in \mathfrak{B} and its corresponding policy π is updated on each iteration step. By [11], the iteration would eventually converge to satisfy the ε -optimal stopping criterion:

$$\left| \max_i (v_i^t) - \min_j (v_j^t) \right| < \varepsilon. \quad (18)$$

when (18) holds, the transition matrix A^{π^*} in \mathfrak{B} and its corresponding π^* give a global ε -optimal policy. By taking limit $\varepsilon \rightarrow 0$, the global optimum of (6) is attained.

IV. GENERALIZED MONOTONIC STRUCTURE

As mentioned in subsection III-A, an efficient algorithm that solves problem (9) without iteration is proposed in this section, and its corresponding time complexity analysis is given in section IV-B. A special pattern of the optimal decision matrix $\mathbf{D}^{b,\pi}$ is revealed by theorem 1, which we name *generalized monotonicity*.

A. Solution of Optimization without Iteration

Intuitively the *generalized monotonicity* captures the feature that the optimal decision matrix has all its non-zero entries lying in a stripe expanding from the top-right corner to the bottom-left. If a entry is non-zero, then the block adjacent to its bottom-right corner should be all-zero. We formally describe and prove it as the following theorem:

Theorem 1 *If \mathbf{D}^{b,π^*} is a solution of (9), and there exist x^-, b^- s.t. $D_{x^-,b^-}^{b,\pi^*} > 0$, then $D_{x^+,b^+}^{b,\pi^*} = 0$ for $\forall x^+ \in \{x \in \mathbb{N} : x^- < x \leq X\}, b^+ \in \{b \in \mathbb{N} : b^- < b \leq B\}$.*

Proof: Suppose there exist x^-, b^- s.t. $0 \leq x^- < X, 0 \leq b^- < B$ and $D_{x^-,b^-}^{b,\pi^*} > 0$ (if not, the case is trivial and Theorem 1 still holds). We prove by contradiction, starting by assuming that $\exists x^+, b^+ \in \mathbb{N}$ s.t. $x^- < x^+ \leq X, b^- < b^+ \leq B$ and $D_{x^+,b^+}^{b,\pi^*} > 0$. Now we pick any positive number δ that satisfies $p_{b^+} + \delta < D_{x^-,b^-}^{b,\pi^*}$ and $p_{b^-} - \delta < D_{x^+,b^+}^{b,\pi^*}$. We formulate another decision matrix $\mathbf{D}^{b,\pi'}$ which is identical to \mathbf{D}^{b,π^*} except the following elements

$$D_{x^-,b^-}^{b,\pi'} = D_{x^-,b^-}^{b,\pi^*} - p_{b^+} + \delta \quad (19.a)$$

$$D_{x^-,b^+}^{b,\pi'} = D_{x^-,b^+}^{b,\pi^*} + p_{b^+} + \delta \quad (19.b)$$

$$D_{x^+,b^+}^{b,\pi'} = D_{x^+,b^+}^{b,\pi^*} - p_{b^-} - \delta \quad (19.c)$$

$$D_{x^+,b^-}^{b,\pi'} = D_{x^+,b^-}^{b,\pi^*} + p_{b^-} - \delta \quad (19.d)$$

because \mathbf{D}^{b,π^*} is a solution of (9), it is easy to verify that $\mathbf{D}^{b,\pi'}$ satisfies (9.b) and (9.c). Also, (9.d) and (9.e) are naturally satisfied by the choice of δ . Hence, $\mathbf{D}^{b,\pi'}$ is a feasible point of (9). Further we have

$$\begin{aligned} \omega_b^{\pi^*} - \omega_b^{\pi'} &= \eta^{-b} \Phi^{X^\top} \text{diag}(\mathbf{p})(\mathbf{D}^{b,\pi^*} - \mathbf{D}^{b,\pi'}) \Phi^B \\ &= \eta^{-b+x^-+b^-} p_{x^+} p_{x^-} (\eta^{x^+-x^-} - 1)(\eta^{b^+-b^-} - 1) > 0. \end{aligned} \quad (20)$$

$\mathbf{D}^{b,\pi'}$ results in a smaller objective value in (9.a) than \mathbf{D}^{b,π^*} , which contradicts the prerequisite that \mathbf{D}^{b,π^*} is a solution.

Therefore, the original assumption must be false, proving the theorem. ■

Now we formally state the Fast Assignment of State Transition (FAST) algorithm that gives a solution to the problem in (9) without iteration. Validity of the FAST algorithm is given

Algorithm 1 The FAST Algorithm

```

1: Initialization:  $m = 0, n = B, u_0 = p_0, w_B = a_B^{b,\pi}, D^{b,\pi} = 0;$ 
2: if  $u_m < w_n$ , then
3:    $D_{m,n}^{b,\pi} \triangleq \frac{u_m}{p_m}, w_n \triangleq w_n - u_m, m \triangleq m + 1, u_m \triangleq p_m,$ 
   go to 9;
4: end if
5: if  $u_m > w_n$ , then
6:    $D_{m,n}^{b,\pi} \triangleq \frac{w_n}{p_m}, u_m \triangleq u_m - w_n, n \triangleq n - 1, w_n \triangleq a_n^{b,\pi},$ 
   go to 9;
7: end if
8:  $D_{m,n}^{b,\pi} \triangleq \frac{w_n}{p_m}, m \triangleq m + 1, n \triangleq n - 1, u_m \triangleq p_m, w_n \triangleq a_n^{b,\pi};$ 
9: if  $m \leq X$  and  $n \geq 0$ , then
10:   go to 2;
11: end if
12:  $D^{b,\pi^*} \triangleq D^{b,\pi};$ 

```

by the following Theorem 2.

Theorem 2 Let D^{b,π^*} be a solution of (9). $\forall m, n \in \mathbb{N}$ s.t. $0 \leq m \leq X, 0 \leq n \leq B$, if denote

$$u_m = (a_n^{b,\pi} - \sum_{i=0}^{m-1} D_{i,n}^{b,\pi^*} p_i) / p_m \quad (21.a)$$

$$w_n = 1 - \sum_{i=n+1}^B D_{m,i}^{b,\pi^*}, \quad (21.b)$$

then $D_{m,n}^{b,\pi^*} = \min(u_m, w_n)$.

Proof: Let $m, n \in \mathbb{N}$ s.t. $0 \leq m \leq X, 0 \leq n \leq B$. By (9.b) to (9.d), we have

$$\sum_{i=0}^m D_{i,n}^{b,\pi^*} p_i \leq a_n^{b,\pi} \quad (22.a)$$

$$\sum_{i=n}^B D_{m,i}^{b,\pi^*} \leq 1 \quad (22.b)$$

the proof is given by contradiction. Assume that neither equation in (22.a) and (22.b) holds. By (9.b) $\sum_{i=0}^X D_{i,n}^{b,\pi^*} p_i = a_n^{b,\pi}$, and also we have $\sum_{i=0}^m D_{i,n}^{b,\pi^*} p_i < a_n^{b,\pi}$, Therefore

$$\exists m^+ > m \quad \text{s.t.} \quad D_{m^+,n}^{b,\pi^*} > 0,$$

by (9.c) and the assumption that the equation in (22.b) does not hold, we have $\exists n^- < n$ s.t. $D_{m,n^-}^{b,\pi^*} > 0$. Therefore, we have $D_{m,n^-}^{b,\pi^*} > 0$ and $D_{m^+,n}^{b,\pi^*} > 0$, contradicting Theorem 1.

Thus, the original assumption is false, i.e., at least one of the equations in (22.a)(22.b) holds. Rewrite (22.a) and (22.b) as

$$D_{m,n}^{b,\pi^*} \leq (a_n^{b,\pi} - \sum_{i=0}^{m-1} D_{i,n}^{b,\pi^*} p_i) / p_m \quad (23.a)$$

$$D_{m,n}^{b,\pi^*} \leq 1 - \sum_{i=n+1}^B D_{m,i}^{b,\pi^*}, \quad (23.b)$$

it is clear that the equation with smaller value on the right side should hold, hence proving theorem 2. ■

B. Complexity Analysis

Suppose the required precision of value iteration is ϵ_v , and interior-point method is applied for all convex optimizations. By [11], the number of iteration steps is bounded by $-C_1 \log \epsilon_v$, where C_1 is a constant. $B+1$ convex optimizations are solved with interior-point method in each iteration step. If precision ϵ_i is required for interior-point method, the number of times Newton's Method is called in each interior-point method is bounded by $-\sqrt{2(B+1)} \log \epsilon_i$ [12]. In Newton's Method the FAST algorithm, whose time complexity is $O(X+B)$, is called $B+1$ times. Therefore, the time cost of value iteration in \mathfrak{B} with the FAST algorithm is bounded by $O((X+B)B^{2.5} \log \epsilon_i \log \epsilon_v)$.

With similar analysis, we have the time complexity of value iteration in \mathcal{S} which is $O(X^2 B^{3.5} \log \epsilon_i \log \epsilon_v)$. An interesting insight into the two complexities is that when $X \ll B$ the difference between value iterations in \mathfrak{B} and \mathcal{S} fades out, which is the case that buffer size is much larger than expected data request in a timeslot. Since there is nearly unlimited caching space, an optimal policy becomes meaningless. However, in most cases, value iteration in \mathfrak{B} with the FAST algorithm brings significant improvement to time efficiency.

V. SIMULATION RESULTS

We first demonstrate the reasonably small compromise of the causal MDP method proposed in this paper, in comparison with the non-causal *tightest string method* which attains the absolute optimum [9]. Then we compare the time consumption of value iteration in \mathfrak{B} , as proposed in this paper, and the conventional MDP method of value iteration in \mathcal{S} . We first assume that data request yields discrete uniform distribution on all integers in $[0, 20]$, and buffer size varies. η takes typical values 1.4 and 2 respectively. Fig. 4 shows that the causal policy designed by MDP compromises little compared with the non-causal optimum with full RDI. The curve *No buffer* corresponds to real-time transmission, and *Infinite buffer* shows the energy cost for stationary transmission at the average rate. The two curves together give the upper and lower bounds of energy cost.

We then fix the buffer size and study the tendency of energy cost as data request pattern changes. Assume that data request always yields discrete uniform distribution on all integers in $[0, X]$. As X grows, Fig. 5 shows that energy cost grows exponentially. The optimal MDP policy still brings significant

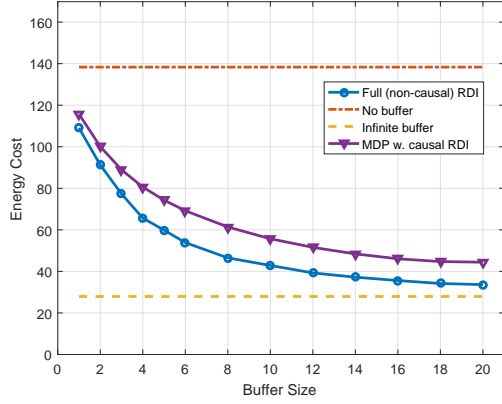


Fig. 4. Average energy cost decreases as buffer size grows, and the causal MDP method attain similar performance to the non-causal optimum. Here, $x \sim U^d[0, 20]$ and $\eta = 1.4$.

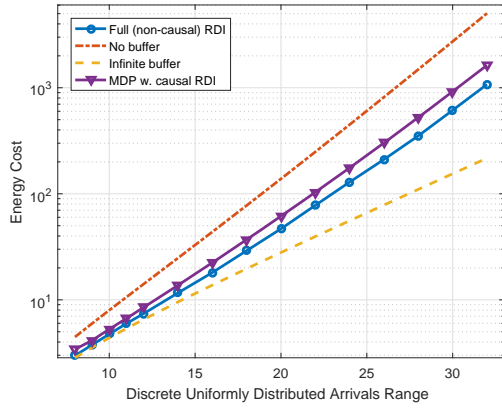


Fig. 5. Average energy cost increases rapidly as data request grows. Similar to the non-causal optimum, causal MDP policy brings significant improvement. Here, buffer size $B = 8$ in content items and $\eta = 1.4$.

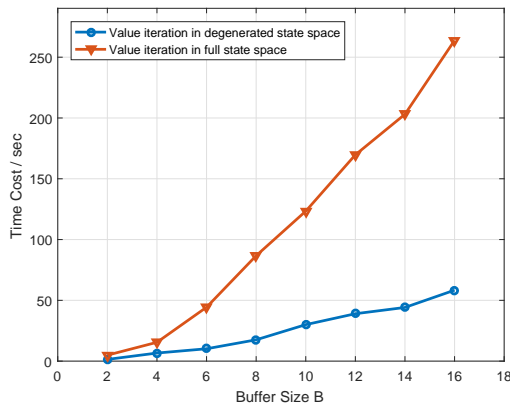


Fig. 6. Value iteration in degenerated space \mathfrak{B} with the FAST algorithm significantly saves time compared with that in \mathcal{S} . Here, the user request is of size $X = 1.5B$ where B denotes the buffer size with $\eta = 1.4$

improvement, and its performance synchronically grows with the non-causal full-RDI method.

As the simulations are done with identical parameter settings, where $X = 1.5B$ always holds and B takes $\{2, 4, \dots, 16\}$, Fig. 6 shows that value iteration in full state space \mathcal{S} consumes much more time than the method proposed in this paper, which is value iteration in \mathfrak{B} with the FAST algorithm, though the two methods end with the same policy.

VI. CONCLUSION

In this paper, we introduced Markov decision processes to the point-to-point proactive caching problem in wireless communications. Though it is possible to directly apply conventional MDP algorithms to design optimal JPC policies, we largely adapted MDP model for the specified problem setting. We revealed and exploited benefits from a special structure, which we proposed as generalized monotonicity. The algorithm designed based on the very structure has significantly accelerated the conventional MDP algorithm in this problem. Additionally, the energy performance of the attained causal policy is comparatively satisfactory, with regard to the globally optimal non-causal policy. Future works may continue to analytically study the improvement of energy consumption, and generalize conclusions in this paper to the scenario of multicasting.

REFERENCES

- [1] "Cisco visual networking index: Forecast and methodology, 2016-2021." [online] available: <https://goo.gl/w4MTvu>.
- [2] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 497–508, April 2012.
- [3] J. G. Andrews, "Seven ways that hetnets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136–144, March 2013.
- [4] U. Niesen, D. Shah, and G. W. Wornell, "Caching in wireless networks," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6524–6540, Oct 2012.
- [5] Z. Chang, Y. Gu, Z. Han, X. Chen, and T. Ristaniemi, "Context-aware data caching for 5g heterogeneous small cells networks," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.
- [6] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.
- [7] M. Gregori, J. Gmez-Vilardeb, J. Matamoros, and D. Gndz, "Wireless content caching for small cell and d2d networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1222–1234, May 2016.
- [8] W. Chen and H. V. Poor, "Content pushing with request delay information," *IEEE Transactions on Communications*, vol. PP, no. 99, pp. 1–1, 2017.
- [9] M. A. Zafer and E. Modiano, "A calculus approach to energy-efficient data transmission with quality-of-service constraints," *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 898–911, Jun. 2009.
- [10] A. C. Gngr and D. Gndz, "Proactive wireless caching at mobile user devices for energy efficiency," in *2015 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2015, pp. 186–190.
- [11] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [12] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.