

Detecting State Transitions of a Markov Source: Sampling Frequency and Age Trade-off

Jaya Prakash Champati, *Member, IEEE*, Mikael Skoglund, *Fellow, IEEE*, Magnus Jansson, *Senior Member, IEEE*,
and James Gross *Senior Member, IEEE*,

Abstract—We consider a finite-state Discrete-Time Markov Chain (DTMC) source that can be sampled for detecting the events when the DTMC transits to a new state. Our goal is to study the trade-off between sampling frequency and staleness in detecting the events. We argue that, for the problem at hand, using Age of Information (AoI) for quantifying the staleness of a sample is conservative and therefore, study another freshness metric *age penalty*, which is defined as the time elapsed since the first transition out of the most recently observed state. We study two optimization problems: minimize average age penalty subject to an average sampling frequency constraint, and minimize average sampling frequency subject to an average age penalty constraint; both are Constrained Markov Decision Problems. We solve them using the Lagrangian MDP approach, where we also provide structural results that reduce the search space. Our numerical results demonstrate that the computed Markov policies not only outperform optimal periodic sampling policies, but also achieve sampling frequencies close to or lower than that of an optimal clairvoyant (non-causal) sampling policy, if a small age penalty is allowed.

Index Terms—Age of information; age penalty; sampling; DTMC source; CMDP;

I. INTRODUCTION

Detecting the occurrence of an event when monitoring an information source or a process of interest is essential to applications from varied domains that include control and information systems. In a control system, for instance, a sensor samples a process for detecting an event where the state of the process exceeds a certain threshold value. In the World Wide Web, a web crawling application is equipped with the task of downloading remote web pages to a local database (for page ranking/indexing etc.), and is required to detect the events when the remote web page gets updated.

In practice, it is impossible to know the exact time instant of occurrence of an event unless the source is sampled infinitely often (or in every time slot for discrete-time systems). However, sampling at a higher frequency incurs costs to a system in terms of the energy consumption of a sensor, or the bandwidth usage of the network for transmitting the samples. On the other hand, sampling at a lower frequency results in staleness in detecting an event. Therefore, we are interested in the question: *given that the source is sampled in time slot n , how to choose the next sampling instant $n + \tau$ such that the conflicting objectives average sampling frequency and average staleness in the event detection are optimized?* In this work,

we address this question for an information source modelled using a finite-state DTMC and the events we want to detect are transitions of the DTMC to new states. Our motivation for studying the DTMC source is that it can serve as an abstract model to describe the states of a plant or a process of interest. For example, we may classify the states of a plant into “good” and “bad” and study a 2-state DTMC by deriving the transition probabilities from the history of observations made on the plant. Another important application domain is maintaining fresh copies of remote sources in a local database [Cho2003]. For example, a web crawler application maintains a local database about web pages on the internet and regularly updates the local database in order to have the information about the latest web pages. It is used, for example, by search engines for page ranking. Here, the time instances of changes of a web page can potentially be modeled using the state transitions in a DTMC. Another merit of studying a DTMC source is that, it allows for tractable analysis for new freshness metrics that take into consideration the source characteristics. This is evident from its use in several related works; for example, see [1]–[4].

The first step in studying the trade-off between sampling frequency and staleness is to choose an appropriate metric for quantifying the staleness of a sample. For this purpose, one may choose Age of Information (AoI), which has emerged as a relevant performance metric for quantifying staleness of updates at a destination in a communication system. It is defined as the time elapsed since the generation of the freshest update available at the destination [5]. However, using the example that follows, we argue that using AoI is conservative for the problem at hand. Consider a two-state DTMC source with states $\{1, 2\}$. The transition probability from state 1 to state 2 is 0.05 and the transition probability from state 2 to state 1 is 0.95. The stationary probabilities for states 1 and 2 are 0.95 and 0.05, respectively. Note that, for a sampling policy to satisfy a unit constraint on AoI, it has to sample the DTMC every two slots, because AoI increases linearly irrespective of the DTMC state that was observed. However, when in state 1, a transition occurs every twenty slots, on average. Thus, sampling every two slots results in redundant samples as each of those samples most likely contain state 1. Thus, in this work we study another staleness metric *age penalty*, which is defined as the time elapsed since the first transition out of the most recently observed state. Note that the age penalty takes into account the current state of the source. As we will see later in Example 2 in Section V, for the above two-state DTMC, using a unit constraint on the age

Jaya P. Champati is with IMDEA Networks Institute, Madrid, Spain. Mikael Skoglund, Magnus Jansson, and James Gross are with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. E-mail: jaya.champati@imdea.org, {skoglund, janssonm, jamesgr}@kth.se. This research was supported in part by the Swedish Research Council (VR) under grant 2016-04404 and in part by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme, grant agreement No. 742648.

penalty results in a sampling policy that samples the DTMC in the next time slot if the observed state in the current time slot is 2, but if the observed state is 1, the policy samples after 7.36 slots, on average. The average sampling frequency of this policy is 0.14. Note the stark contrast to the policy, with sampling frequency 0.5, resulting from using AoI; this signifies the importance of using age penalty. It is worth noting that age penalty is an instantiation of a general metric Age of Incorrect Information (AoII) proposed in [4].

We formulate two problems: minimize average age penalty subject to an average sampling frequency constraint, and minimize average sampling frequency subject to an average age penalty constraint. Both the problems are Constrained Markov Decision Problems (CMDPs). We then formulate the Lagrangian MDP with parameter λ for which we prove two structural results: 1) the optimal λ belongs to the interval $(0, 1)$, and 2) given a state i is observed at a sampling instant, the optimal sampling interval for the next sample is upper bounded by $(1-\lambda)/\log p_{ii}+1$, where p_{ii} is the self-state transition probability in state i . These two structural results reduce the search space significantly and we use the Relative Value Iteration (RVI) algorithm (cf. [6]) to compute the optimal deterministic policies for the Lagrangian MDP. To compute an optimal policy for the CMDP, we use a classical result that the optimal Markov policy for a single constraint CMDP can be obtained by randomizing between two deterministic optimal policies for the Lagrangian MDPs [7].

In our numerical analysis, we find that the optimal policy always provides lower sampling frequency than the best periodic sampling policy and the gap increases with lower probability of transitions to other states. We also present a comparison of the sampling frequency achieved by the optimal policy with that of the sampling frequency of an optimal clairvoyant (non-causal) sampling policy. Interestingly, if the system can allow a small age penalty equal to one slot, then one can achieve much lower sampling frequencies than that of the optimal clairvoyant sampling policy in some scenarios.

The rest of the paper is organized as follows. Related work is presented in Section II. In Section III, we present the system model and the CMDPs. The Lagrangian MDP solution approach and the structural results are presented in Section IV. Numerical results are presented in Section V, and we conclude in Section VI.

II. RELATED WORKS

In the AoI literature, the works [1]–[3], [8]–[11] considered remote monitoring/estimation of the states of a Markov source. The authors in [8] proposed a freshness metric based on the mutual information between the current state of the source and the received states at a remote monitor, and solved an optimal sampling problem for maximizing the mutual information. In [9], the authors analysed freshness by proposing a closely related metric to age penalty based on conditional entropy, where the current state and the states in the past till the generation time of the freshest sample at the monitor are conditioned with respect to this freshest sample. The authors also studied the detection delay of each state change and focused on the

analysis of these metrics, but focused on studying periodic sampling and zero-wait policies using simulation. We note that in contrast to the detection delay, age penalty only considers the first state change since the last sampling instant that allows tractability while capturing the staleness.

Displaying the state of a continuous-time Markov chain source at a remote monitor was studied in [10]. The authors analysed the probability of error in displaying the correct state of the source. In our system model, we consider staleness only at the sampler. In [4], the authors proposed the AoII metric, which is defined as a product of two functions: one is any increasing function of time until the time the next update is received, and the other function is a general penalty function for mismatch in the states at the source and the receiver. The authors studied an instance of the AoII metric that is a linearly increasing function with time whenever the freshest state at the receiver and the DTMC are unequal. They study a parameterized symmetric N -state DTMC, where the transition probability from a state to itself is p_t and the probability of transition from a state to any other state is p_R . As a consequence, for the , these values has to satisfy $p_R + (N - 1)p_t = 1$. In [1], the authors considered remote monitoring of a two-state Markov Chain with geometrically distributed communication delays and assumed that the sampler knows the state in each slot. They computed the average AoI and the estimation error for two sampling policies: 1) zero-wait policy, which sends the latest sample when the channel is idle, and 2) sample-at-change policy, which sends the latest sample when the channel is idle and a transition to a state different from the previous sample occurs. Further, for this system, the authors computed optimal sampling policies for three different performance metrics, namely, the estimation error, AoI, and AoII [2]. In [3], the authors considered an N -state DTMC where transitions occur only between adjacent states, and geometrically distributed communication delays. They minimize an instantiation of AoII – the absolute difference between the states at the receiver and the Markov source multiplied by the current time – subject to a constraint on the power associated with the transmission attempts. In contrast to the above works, we do not consider communication delay, but study the general N -state ergodic DTMC without any further assumptions on the structure of the state transition probability matrix. Furthermore, we consider that the sampler only knows the state of the DTMC when it samples. In [11], the authors considered multiple Markov sources and their metric is an instantiation of AoII called Mean Age of Incorrect Information (MAoII), which penalizes linearly all the time slots where the freshest state at the receiver and the DTMC are unequal. They also studied the structured transition probability considered in [4], and proposed a Whittle-Index policy to solve the minimization of the sum of MAoII for the sources subject to a communication constraint. Finally, several works studied AoI optimization under a constraint for different system settings, for example cf. [12]–[16], and formulated CMDPs and used the Lagrangian method to solve it. For the CMDP at hand, we provide novel structural results that reduce the runtime complexity of solving for the optimal policy and also prove the tightness of the constraint under the optimal policy.

The problem of when to sample next has been studied for many years in control theory, see for example [17]–[20]. In [17], [18], the authors considered the off-line (on-line) problem of choosing the time instants to sample sensor measurements to minimize a Linear Quadratic Gaussian (LQG) cost in a Linear Time Invariant (LTI) system. In [19], the authors considered minimizing squared error distortion for state estimation of a Markov source under a constraint on the maximum number of transmitted samples. We note however that, in this work the sensor is assumed to sample the process continuously but only transmits certain samples based on a criterion (event-triggering). In [20], the authors studied the design of sampling intervals such that the stability of a non-linear stochastic dynamical system is ensured. In all the above works, the objective is either to minimize estimation error or control cost or ensure stability of the system.

Perhaps the most relevant application for the problem we have studied is maintaining fresh local database of remote sources which is essential in systems like web crawling [21], [22]. The authors in [21] proposed the following *age* metric: at any given time, the age of a webpage is the time elapsed since the webpage was first modified since last synchronization. Note that the age penalty metric we study is similar to the age metric proposed in [21] with a subtle difference that the latter metric is a function of time, whereas the age penalty is assigned only when there is a sample. A static optimization problem was solved in [21] by modeling the webpage modifications with a Poisson process and optimal periodic policies were computed for minimizing the age metric. However, we note that dynamic policies that use the state of the system have not been studied in this line of work; see [22] for a survey. In contrast, we considered the general set of causal sampling policies and studied the trade-off between sampling frequency and age penalty for detecting state transitions in a finite-state DTMC.

III. SYSTEM MODEL AND PROBLEM STATEMENT

A. Markov Source

We consider an information source/process that is modelled by an N -state DTMC $\{X_n, n \geq 0\}$ where $N < \infty$. We assume that the DTMC is ergodic, i.e., it is irreducible [23]. Let $S = \{1, 2, \dots, N\}$ denote the set of states. Let $P = [p_{ij}]$, $i, j \in S$ denote the transition probability matrix. Here, p_{ij} is the one-step probability of the state transitioning from state j to state i . Further, let $p_{ij}^{(n)}$ denote the n -step transition probabilities. We have

$$p_{ij}^{(n)} = \mathbb{P}(X_n = i | X_0 = j), \forall i, j \in S.$$

Let $\xi = [\xi_1, \xi_2, \dots, \xi_N]$ denote the steady-state vector, where ξ_j denotes the steady-state probability of finding the DTMC in state j . Since the DTMC is ergodic, $\xi_j > 0$, for all j .

A time slot in the system represents one unit of time of the DTMC and the state transitions occur at the start of a time slot. The state of the DTMC can only be observed by sampling the source; see Figure 1. Let $T_0 = 0, T_1, T_2, \dots$ denote the time instants of new-state transitions of the DTMC. We are interested in detecting these transitions at the earliest

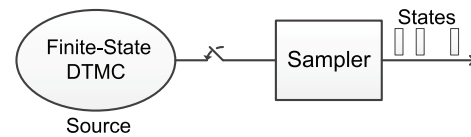


Fig. 1: Sampling an information source/process modelled using a DTMC.

time possible. This problem has relevance to applications from different domains:

- In a control system, where monitor/controller is co-located with the sampler, the source is a process of interest, and a state transition could represent an event that the process signal exceeds a certain threshold.
- Another application domain is information systems which maintain fresh copies of remote sources in a local database [Olsten2010]. For example, a web crawler application maintains a local database about web pages on the internet and regularly updates the database in order to have the information about the latest web pages. The database is used, for example, by search engines for page ranking. In this example, the download time of the updates of a webpage is relatively negligible compared with the frequency of polling and the changes in the webpages are the state transitions.

Clearly, sampling the source at the start of every slot allows us to detect each and every transition of the DTMC. Instead, our aim here is to use a lower sampling frequency. This translates to energy¹ savings for a sensor and/or bandwidth savings for transmitting lower number samples to a controller/monitor. In the case of a web crawling application, this translates to a lower frequency of downloads of the remote web page. However, using lower sampling frequency will result in staleness in detecting a transition and may also lead to missing several transitions. We are thus interested in studying the trade-off between sampling frequency and staleness. Next, we define sampling policies and the age penalty for quantifying staleness.

B. Sampling Policy and Age Penalty

Assume that X_0 is given. A sampling policy π specifies the set of sampling instants $\{G_k, k \geq 1\}$, where G_k is the sampling instant of the k th sample. Define $\tau_k = G_k - G_{k-1}$ for all $k \geq 1$, and $G_0 = 0$, then the policy π can be equivalently specified by $\{\tau_k, k \geq 1\}$. We assume that $\tau \in Q = \{1, 2, \dots, M\}$, where M is the maximum inter-sampling time allowed in the system. We will see later that the proposed solution approach can accommodate the case where Q is countably infinite. Let Π denote the set of all causal policies, where a causal policy considers the current and all past observed states and past actions for choosing the current action. In the sequel, we study the following policies.

¹Explicit modeling of energy consumption in the system could involve more system design choices apart from sampling frequency and we do not pursue it in this work as it complicates analyses and obfuscates key insights into the problem.

- 1) *Markov policies*: A Markov policy maps each state to an action with a fixed probability. To be precise, let j be the observed state in the k th decision epoch, then, under a Markov policy, τ_k is assigned a value $\tau \in \mathcal{Q}$ according to a fixed probability distribution $\mathbb{P}^\pi(\tau_k = \tau | j)$. Let Π^{MR} denote the set of Markov policies.
- 2) *Deterministic policies*: A deterministic policy is a Markov policy that maps each state to some action with probability 1. Let Π^{D} denote the set of deterministic policies.
- 3) *Periodic sampling policies*: Under these policies, samples are taken at fixed time intervals τ . With a slight abuse in notation we use $\pi(\tau)$ to denote such a policy. Note that periodic sampling policies are a subclass of Markov policies.
- 4) *Optimal clairvoyant sampling policy*: Under this policy, the next transition to a new state is assumed to be known a priori, and thus the source is sampled exactly at the instants when transitions between states occur. Let $\pi^\dagger = \{G_k^\dagger, k \geq 1\}$ denote this policy and ν^\dagger denote its average sampling frequency. Note that π^\dagger is a non-causal policy and we study it for theoretical bench-marking.

As stated before, sampling the source at the start of every slot allows us to identify each and every transition of the DTMC to a new state and thus staleness of each sample is zero. However, quantifying the staleness of a sample in general is not entirely obvious. This is because, when the sampler samples the source it may find that the DTMC is in the same state or a different state from the previous sample, and even in the former case multiple transitions might have occurred. One may consider AoI, denoted by $\Delta(t)$, at the sampler as the staleness metric. It increases linearly between two sampling instants and resets to zero at the sampling instants. However, using this staleness metric is conservative in this context. To illustrate this, in Figure 2 we plot the sample-path of a 3-state DTMC and the resulting AoI. Note that in the duration between the instants G_1 and G_2 , the DTMC stays in state 2 for 3 time-slots after it was observed by the sampler at G_1 . Ideally, this should not be accounted for the staleness of the sample at G_2 , but AoI adds a linear penalty for this duration. Ideally, this should not be accounted for the staleness of the sample at G_2 , but AoI adds a linear penalty for this duration.

Using the above insight, we quantify the staleness of a sample k by introducing *age penalty* A_k , which is defined as the time elapsed since the first transition out of the state in the $k-1$ sample². Under policy π the age penalty for the k th sample is given by

$$A_k(\pi) = \max\{0, G_k - \min_n \{T_n : T_n \geq G_{k-1}\}\}.$$

This metric is illustrated and contrasted with AoI in Figure 2. Note that, as the sampling period increases the sampling instant G_k increases and by definition the age penalty increases. Under a policy π , the average age penalty, denoted by $\mathbb{E}[A(\pi)]$, is given by,

$$\mathbb{E}[A(\pi)] = \limsup_{K \rightarrow \infty} \frac{\mathbb{E}[\sum_{k=1}^K A_k(\pi)]}{K},$$

²One may also consider including the number of missed transitions in the age penalty and with some effort solve the problem using the same approach as in this paper.

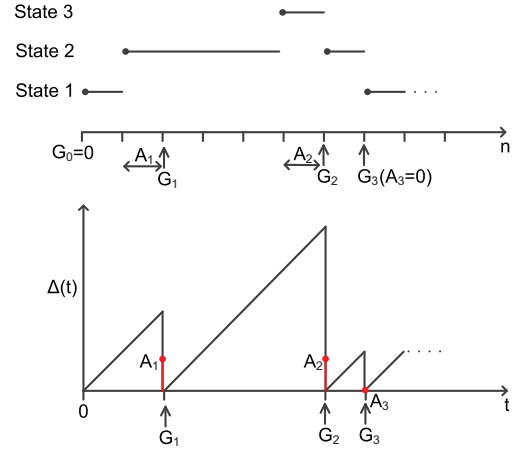


Fig. 2: A sample path of a 3-state Markov chain. AoI and age penalties are depicted for the first three sampling instants of a policy with $G_1 = 2, G_2 = 6$, and $G_3 = 7$.

and the average sampling-interval is given by $\limsup_{K \rightarrow \infty} \frac{\mathbb{E}[\sum_{k=1}^K \tau_k]}{K}$, where the expectation is with respect to the probability distribution induced by π on the observed states and the actions.

C. Optimization problems \mathcal{P}_1 and \mathcal{P}_2

We are interested in the following two problems. In problem \mathcal{P}_1 , we minimize the average age penalty for a given upper bound $\nu \in (0, 1)$ on the average sampling frequency:

$$\begin{aligned} \mathcal{P}_1 : & \text{minimize}_{\pi \in \Pi} \mathbb{E}[A(\pi)] \\ & \text{s.t.} \quad \limsup_{K \rightarrow \infty} \frac{\mathbb{E}^\pi[\sum_{k=1}^K \tau_k]}{K} \geq \frac{1}{\nu}. \end{aligned} \quad (1)$$

In problem \mathcal{P}_2 , we maximize the average sampling-interval for a given upper bound $d \geq 0$ on the average age penalty:

$$\begin{aligned} \mathcal{P}_2 : & \text{maximize}_{\pi \in \Pi} \limsup_{K \rightarrow \infty} \frac{\mathbb{E}^\pi[\sum_{k=1}^K \tau_k]}{K} \\ & \text{s.t.} \quad \mathbb{E}[A(\pi)] \leq d. \end{aligned} \quad (2)$$

Let π_1^* and π_2^* denote optimal policies for \mathcal{P}_1 and \mathcal{P}_2 , respectively.

Remark 1: For \mathcal{P}_1 , an optimal periodic sampling policy chooses $\tau = \lceil 1/\nu \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. To see this, under the periodic sampling policy with sampling period one time slot, the age penalty will be zero in each slot, because in this case the sampler will know the exact state of the DTMC in all time slots. As the sampling period increases, the average age penalty increases monotonically. Therefore, $\lceil 1/\nu \rceil$ is the minimum period for a periodic sampler that has the least age penalty while satisfying the lower bound constraint $1/\nu$. For \mathcal{P}_2 , an optimal periodic sampling policy chooses $\tau = d + 1$. To see this, a periodic sampler with sampling period a has the average age penalty at most $a - 1$. Therefore, the periodic sampling policy with sampling period $d + 1$ has the maximum sampling period among all periodic sampling policies that satisfy the average age penalty constraint d .

Finally, we define $\tau^\dagger = \lceil 1/\nu^\dagger \rceil$, where ν^\dagger is the average clairvoyant sampling frequency.

IV. SOLUTION APPROACH

Both \mathcal{P}_1 and \mathcal{P}_2 are Constrained Markov Decision Problems (CMDP) with finite state and action sets and therefore, an optimal policy exists and it belongs to the set of Markov policies [cf. Theorem 2.1 [7]]. Thus, in the following we only need to consider the set of Markov policies. Under Markov policies the induced stochastic process $\{X_{G_k}, k \geq 1\}$, i.e., the sequence of observed states, is also a DTMC; in the sequel we refer to it as the *induced DTMC*.

Two main approaches to solve a CMDP with finite-state and finite-action sets are: 1) using Linear Programming (LP) [24], and 2) using Lagrangian MDP [25], [26]. In [27], we used the LP approach and solved the resulting linear programs using MATLAB. However, we note that these linear programs – the formulations are given in Appendix B – have MN number of variables, and the computational complexity $\text{cmp}(LP)$ of the practical state-of-the-art algorithms³ is a function of $(MN)^{2.8}$. Therefore, the run-time of the LP approach does not scale well if M and N are large. In the following, we use the Lagrangian MDP approach to solve the CMDP and also provide a justification for a reduction in the computational complexity of the proposed method. The steps in this approach are summarized below:

- 1) We formulate the CMDP and the corresponding Lagrangian MDP with parameter $\lambda \geq 0$. Since the Lagrangian MDP is an unconstrained MDP, an optimal deterministic policy can be computed by solving Bellman's equations.
- 2) We prove structural results for the Lagrangian MDP that reduce the search space (making it independent of M), and use them to propose an efficient algorithm for computing the optimal policies.
- 3) Given two optimal policies, corresponding to two chosen λ values, for the Lagrangian MDP, we compute the optimal Markov policy using randomization [25], [26].

In the sequel, we describe the solution approach for \mathcal{P}_1 and the same procedure applies to \mathcal{P}_2 .

A. Elements of the CMDP

The decision epochs in the CMDP are indexed by k .

- **State space:** $S = \{1, 2, \dots, N\}$.
- **Action space:** At decision epoch k , the next inter-sampling time τ_{k+1} is chosen from the set $Q = \{1, 2, \dots, M\}$.
- **Transition probabilities:** The next state $j \in S$ of the induced DTMC only depends on the current observed/sampled state i and the sampling interval τ . To be precise, let i be the state observed in decision epoch k , i.e., in time slot G_k , then the transition probability of the induced DTMC to state j for any sampling interval τ is given by

$$\begin{aligned} q_{i\tau j} &= \mathbb{P}(X_{G_k+\tau} = j | X_{G_k} = i) \\ &= \mathbb{P}(X_\tau = j | X_0 = i) = p_{ij}^{(\tau)}, \forall i, j \in S \text{ and } \tau \in Q. \end{aligned}$$

³Considering that a recent algorithm from [28] is used for solving the linear programs, while Strassen's algorithm is used for matrix multiplications.

Further, given $\pi \in \Pi^{\text{MR}}$, the steady-state probabilities $\mathbb{P}^\pi(X_{G_k} = j)$ for the induced DTMC can be computed from the following transition probabilities:

$$\begin{aligned} \mathbb{P}(X_{G_{k+1}} = j | X_{G_k} = i) &= \mathbb{E}[q_{i\tau j}] \\ &= \sum_{\tau=1}^M q_{i\tau j} \mathbb{P}^\pi(\tau | i), \forall i, j \in S. \end{aligned} \quad (3)$$

- **Costs:** In decision epoch k , if the observed state $X_{G_k} = j$, then choosing the next sampling interval $\tau_{k+1} = \tau \in Q$ results in a cost contributing to the average age-penalty, given by $c_k(X_{G_k} = j, \tau_{k+1} = \tau)$. Note that the first transition of the DTMC to a new state from state $X_{G_k} = j$ could occur in any of the τ slots with probability p_{jj} . Consider that this first transition occurred in the n th slot from G_k , then the age penalty incurred in epoch k is equal to $\tau - n$; this event happens with geometric probability $(1 - p_{jj})p_{jj}^{n-1}$. Thus, we obtain

$$\begin{aligned} c_k(X_{G_k} = j, \tau_{k+1} = \tau) &= \sum_{n=1}^{\tau-1} (\tau - n)(1 - p_{jj})p_{jj}^{n-1} \\ &= \sum_{n=1}^{\tau-1} \tau(1 - p_{jj})p_{jj}^{n-1} - \sum_{n=1}^{\tau-1} n(1 - p_{jj})p_{jj}^{n-1} \\ &= \tau(1 - p_{jj}^{\tau-1}) - \frac{(1 - \tau p_{jj}^{\tau-1})(1 - p_{jj}) + p_{jj} - p_{jj}^\tau}{1 - p_{jj}} \\ &= \tau - \frac{1 - p_{jj}^\tau}{1 - p_{jj}}, \end{aligned} \quad (4)$$

and the cost contributing to the average sampling-interval is τ . It is easy to see that

$$\mathbb{E}[A(\pi)] = \lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}^\pi \left[\sum_{k=0}^{K-1} c_k(X_{G_k}, \tau_{k+1}) \right],$$

where the expectation \mathbb{E}^π is with respect to the induced probability distribution under π .

The CMDP for \mathcal{P}_1 is:

$$\begin{aligned} \min_{\pi \in \Pi^{\text{MR}}} \quad & \lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}^\pi \left[\sum_{k=0}^{K-1} c_k(X_{G_k}, \tau_{k+1}) \right] \\ \text{s.t.} \quad & \limsup_{K \rightarrow \infty} \frac{\mathbb{E}[\sum_{k=1}^K \tau_k]}{K} \geq \frac{1}{\nu}. \end{aligned} \quad (5)$$

B. Lagrangian MDP

Since \mathcal{P}_1 has one constraint, an optimal Markov policy can be obtained by randomizing between two deterministic policies, which can be computed using a Lagrangian MDP formulation of (5). In the following, we formulate this Lagrangian MDP, with parameter λ , and compute the deterministic policies and the randomization factor to obtain an optimal Markov policy.

In decision epoch k , if the observed state $X_{G_k} = j$, then choosing the next sampling interval $\tau_{k+1} = \tau \in Q$ results in a Lagrangian per-stage cost given by

$$c_k^\lambda(X_{G_k} = j, \tau_{k+1} = \tau) = c_k(j, \tau) - \lambda\tau.$$

Let

$$\Phi(\lambda, \pi) = \lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}^\pi \left[\sum_{k=0}^{K-1} c_k^\lambda(X_{G_k}, \tau_{k+1}) \right].$$

For $\lambda \geq 0$, the Lagrangian MDP for \mathcal{P}_1 is:

$$\min_{\pi \in \Pi^D} \Phi(\lambda, \pi). \quad (6)$$

Note that in (6), we consider minimization over Π^D because the Lagrangian MDP is an unconstrained MDP with finite states and finite action sets, for which an optimal policy belongs to the set of deterministic policies Π^D . Also, note that we left out the term $\frac{\lambda}{\nu}$ in (6) as it is a constant and does not affect the solution of the Lagrangian MDP. Let π_λ^* denote an optimal policy for (6). Define

$$\phi(\lambda) = \Phi(\lambda, \pi_\lambda^*) + \frac{\lambda}{\nu},$$

then from the Lagrangian duality result [7], we have

$$\mathbb{E}[A(\pi^*)] = \max_{\lambda \geq 0} \phi(\lambda). \quad (7)$$

Also, from Lagrangian duality theory we have that $\phi(\lambda)$ is concave in λ [29].

Lemma 1. *Let λ^* maximize $\phi(\lambda)$, then $0 < \lambda^* < 1$.*

Proof. From (4), we infer that $c_k(X_{G_k}, \tau_{k+1}) < \tau_{k+1}$ for all k . This implies, for all k ,

$$c_k^\lambda(X_{G_k}, \tau_{k+1}) = (1 - \lambda)\tau_{k+1} - \frac{1 - p_{jj}^{\tau_{k+1}}}{1 - p_{jj}} < 0, \text{ if } \lambda \geq 1. \quad (8)$$

From (8), we conclude that if $\lambda \geq 1$, then by choosing $\tau_1 = \infty$ results in $\Phi(\lambda, \{\infty\}) = -\infty$, i.e., a trivial policy which does not sample minimizes $\Phi(\lambda, \pi)$. In other words, for all $\lambda \geq 1$, we have $\pi_\lambda^* = \{\infty\}$ and $\phi(\lambda) = -\infty$. However, from (7), we infer that λ^* cannot be greater than 1 since the optimal expected age penalty in the LHS is positive (since the average frequency constraint $\nu < 1$). On the other hand, if $\lambda = 0$, then choosing $\tau_k = 1$, for all k , results in $c_k^\lambda(X_{G_k}, \tau_{k+1}) = 0$. This implies $\Phi(0, \pi = \{1, 1, \dots\}) = 0$ and we obtain $\phi(0) = 0$. Again, from (7), $\phi(\lambda^*)$ cannot be 0 since the optimal expected age penalty in the LHS is positive, and hence $\lambda^* \neq 0$. \square

Let $\pi(i) \in Q$ denote the action taken by the policy π when in state i .

Lemma 2. *Given $\lambda \in (0, 1)$, under the optimal policy π_λ^* , we have $\pi_\lambda^*(i) < \log(1 - \lambda) / \log p_{ii} + 1$.*

Proof. To prove the result, we first present an equivalent formulation of the CMDP (for \mathcal{P}_1) using a binary action space and redefining the state space. We formulate the corresponding Lagrangian MDP with parameter λ and then prove the result using Bellman's optimality equations for the Lagrangian MDP and the value iteration steps. In particular, we show that, in the first step of the value iteration algorithm if we set the inter-sampling interval less than the value $\log(1 - \lambda) / \log p_{ii} + 1$, then in all the subsequent iterations of the algorithm, the inter-sampling times will be less than $\log(1 - \lambda) / \log p_{ii} + 1$. The detailed proof is deferred to Appendix A. \square

Using Lemma 2, we reduce the search space by defining action sets \hat{Q}_i corresponding to each state i as follows: $\hat{Q}_i = \{1, 2, \dots, \lceil \log(1 - \lambda) / \log p_{ii} \rceil\}$.

Algorithmic steps: We now describe three steps in which we compute the optimal Markov policy. First, given λ , we compute the optimal deterministic policy for the Lagrangian MDP using the Relative Value Iteration (RVI) algorithm (cf. [6]). It solves the following Bellman's equations:

$$h(i) = \min_{\tau \in \hat{Q}_i} \left\{ c_k^\lambda(i, \tau) + \sum_j q_{i\tau j} h(j) \right\} - \gamma, \quad \forall i > 1$$

$$\gamma = \min_{\tau \in \hat{Q}_1} \left\{ c_k^\lambda(1, \tau) + \sum_j q_{1\tau j} h(j) \right\}.$$

Note that we only search over the reduced action sets \hat{Q}_i due to Lemma 2. Given λ , RVI outputs π_λ^* and $\phi(\lambda)$. Using the fact that $\phi(\lambda)$ is concave in λ , in the second step, we use bisection search for λ in the interval $(0, 1)$ and compute two optimal deterministic policies corresponding to two λ values, namely, λ_1 and λ_2 , which belong to the left and right ϵ -neighborhoods of the optimal λ , for some positive ϵ close to zero. The bisection search is presented in Algorithm 1. In line 5 of Algorithm 1, we check the condition for the derivative of $\phi(\lambda)$ given below:

$$\phi'(\lambda) = \lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}^{\pi_{\lambda^*}} \left[\sum_{k=0}^{K-1} (-\tau_{k+1}) \right] + \frac{1}{\nu}.$$

Note that $\phi'(\lambda) > 0$ implies that the sampling interval constraint is violated by π_{λ^*} .

Algorithm 1: Bisection Search

- 1: Initialize $\epsilon > 0$ and $\delta > 0$ (close to zero).
 - 2: $\lambda_1 = \delta$, and $\lambda_2 = 1 - \delta$
 - 3: **while** $\lambda_2 - \lambda_1 > \epsilon$ **do**
 - 4: $\lambda = (\lambda_1 + \lambda_2) / 2$
 - 5: Use RVI to compute $\phi(\lambda)$
 - 6: **If** $\phi'(\lambda) > 0$, **then** $\lambda_1 = \lambda$, **else** $\lambda_2 = \lambda$.
 - 7: **end while**
 - 8: Output $\pi_{\lambda_1}^*$ and $\pi_{\lambda_2}^*$.
-

In the third step, we compute the randomization factor α . Given the deterministic optimal policies $\pi_{\lambda_1}^*$ and $\pi_{\lambda_2}^*$ output by Algorithm 1, an optimal Markov policy for the CMDP can be computed by randomizing between those two policies [26]. However, to compute the randomization factor, we require that the constraint of the CMDP is tight under an optimal Markov policy. This is indeed true for the CMDP at hand and is stated in the following lemma.

Lemma 3. *Under an optimal Markov policy π_1^* that solves \mathcal{P}_1 , the constraint is tight, i.e.,*

$$\limsup_{K \rightarrow \infty} \frac{\mathbb{E}^{\pi_1^*} \left[\sum_{k=1}^K \tau_k \right]}{K} = \frac{1}{\nu}.$$

Proof. For this proof, we use the approach of formulating the CMDP as a linear program (cf. [7]) and show that the constraint in the linear program is tight by using contradiction. The details of the proof are deferred to Appendix B. \square

Given $\pi_{\lambda_1}^*$ and $\pi_{\lambda_2}^*$, and the tightness result, we compute

$$\tau_1 = \limsup_{K \rightarrow \infty} \frac{\mathbb{E}^{\pi_{\lambda_1}^*} [\sum_{k=1}^K \tau_k]}{K},$$

$$\tau_2 = \limsup_{K \rightarrow \infty} \frac{\mathbb{E}^{\pi_{\lambda_2}^*} [\sum_{k=1}^K \tau_k]}{K},$$

and randomization factor

$$\alpha = (1/\nu - \tau_2)/(\tau_1 - \tau_2). \quad (9)$$

Theorem 1. *There exists $\epsilon_0 > 0$ such that for all $\epsilon \leq \epsilon_0$, the following Markov policy is optimal for \mathcal{P}_1 : at the start of the experiment, choose $\pi_{\lambda_1}^*$ with probability α , and choose $\pi_{\lambda_2}^*$ with probability $1 - \alpha$.*

Proof. The result directly follows from the construction of the Lagrangian MDP [7], the Lemmas 1, 2 and 3, and the randomization method in [26]. \square

In Theorem 1, ϵ_0 refers to the maximum difference between λ_1 and λ_2 below which randomizing between the two optimal deterministic policies for the respective Lagrangian MDPs results in an optimal Markov policy for \mathcal{P}_1 .

Remark 2: The computational complexity of Algorithm 1 is given by $O(\text{cmp}(RVI) \log \frac{1}{\epsilon})$, where $\text{cmp}(RVI)$ is the computational complexity of RVI. Note that the computational complexity is independent of M as the search space in RVI iterations is over the sets $\hat{Q}_i = \{1, 2, \dots, \lceil \log(1-\lambda)/\log p_{ii} \rceil\}$. We observed from simulations that Algorithm 1 is much faster, by a factor of 0.25, than using an LP even for small value $M = 20$, and that the run time of LP exponentially increases, as M increases, resulting in orders of magnitude higher run time compared to Algorithm 1.

In Figure 3, we compare the runtimes of Algorithm 1 (including the computation of randomization factor α) and the LP solution used in our workshop paper [27]. The runtimes are computed using the MATLAB function *timeit* on a system with 2.70 GHz CPU and 16 GB RAM. For $N = 10$ and $N = 50$, we use the parameterized transition probability matrix defined in Section IV.B, with parameter $p = 0.9$. Observe that the runtime of Algorithm 1 is lower by a factor of 4 compared to the runtime of the LP solution even for a small value of $M = 50$. Thanks to Lemma 2, the runtime of Algorithm 1 does not depend on M . In contrast, as M increases, the runtime of the LP solution increases exponentially, making its runtimes higher by more than two orders of magnitude at $M = 1000$.

Remark 3: We note that \mathcal{P}_2 can be solved using the Lagrangian MDP approach. In the following, we present a method to solve \mathcal{P}_2 by leveraging the solutions of \mathcal{P}_1 . From Lemma 3, we have that the constraint of \mathcal{P}_1 is tight. Therefore, the inter-sampling time constraint $1/\nu^*$ and the corresponding minimum average age penalty, say d^* , form a Pareto-optimal pair. In other words, if d^* is the age limit in \mathcal{P}_2 , then the maximum inter-sampling time will be $1/\nu^*$. Using this observation, we propose to solve \mathcal{P}_2 using a bisection search and the solution method for \mathcal{P}_1 . Given age limit d in \mathcal{P}_2 , we start with the upper bound $a := 1$ and lower bound $b := 0$ and set $\nu := (a + b)/2$ and solve \mathcal{P}_1 using Algorithm 1 and the randomization factor α . If the minimum age penalty is less

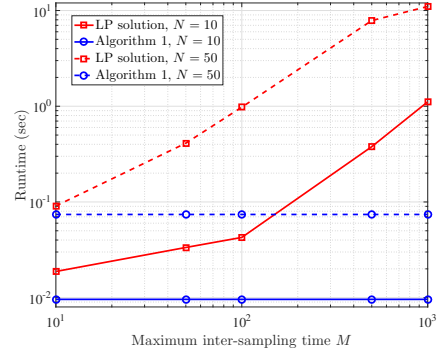


Fig. 3: Runtimes of the solution approaches with varying M , different values of N , and $p = 0.9$.

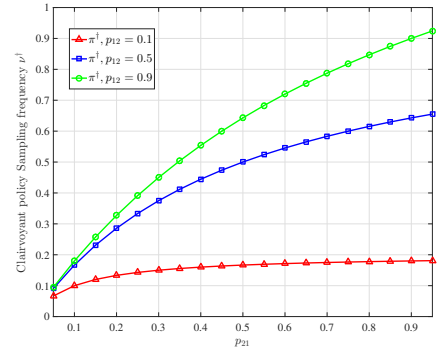


Fig. 4: Sampling frequency under π^\dagger for a two-state Markov chain.

than d , then we set $b := (a + \nu)/2$, else we set $b := (\nu + b)/2$, and we repeat the steps with $\nu := (a + b)/2$. Note that, this algorithm is guaranteed to converge because there exists a maximum inter-sampling interval that can be achieved for a given age limit d and the existence of a Markov policy for \mathcal{P}_2 is guaranteed as it is a CMDP with finite-state and finite-action sets.

C. Computing ν^\dagger

Note that, in \mathcal{P}_1 the value of ν in the constraint can be chosen in the interval $(0, 1]$. We are particularly interested in setting $\nu = \nu^\dagger$, because this will give us the minimum achievable average age-penalty for the same sampling frequency achieved by the optimal clairvoyant sampling policy π^\dagger . We note that ν^\dagger can be obtained by subtracting the percentage of the total frequency of transitions in the DTMC contributed due to self transitions, i.e., transitions from a state to itself, from the total frequency of transitions in the DTMC. The state this result in the following proposition.

Proposition 1. *Under the optimal clairvoyant sampling policy π^\dagger , with probability one, the average sampling frequency ν^\dagger is given by*

$$\nu^\dagger = 1 - \sum_{j=1}^N \xi_j p_{jj}.$$

Proof. Let $Z_j(K)$ denote a random variable that counts the number of self transitions from state j to itself from time

slot zero to time slot K . Let Y_k denote a random variable that equals one if the state of the DTMC is j in slots $k-1$ and k , and equals zero, otherwise. Formally, $Y_k = \mathbb{1}\{X_k = j, X_{k-1} = j\}$, where $\mathbb{1}\{\cdot\}$ is an indicator random variable. By definition, we have

$$\begin{aligned} Z_j(K) &= \sum_{k=1}^K Y_k \\ \implies \lim_{K \rightarrow \infty} \frac{Z_j(K)}{K} &= \lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K Y_k}{K}. \end{aligned} \quad (10)$$

Note that the LHS of the last step above denotes the fraction of slots in which self transitions occur in state j in the steady state. We simplify the RHS as follows. Note that the random variables Y_k and Y_{k-1} are dependent random variables, but Y_k and Y_{k-2} are independent for all $k > 1$. Therefore, we apply the strong law of large numbers (cf. Lemma 2 [30]) and obtain, with probability one,

$$\begin{aligned} \lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K Y_k}{K} &= \mathbf{E}[\mathbb{1}\{X_k = j, X_{k-1} = j\}] \\ &= \mathbf{E}[\mathbb{1}\{X_k = j|X_{k-1} = j\} \mathbb{1}\{X_k = j\}] \\ &= \mathbf{E}[\mathbb{1}\{X_k = j|X_{k-1} = j\}] \mathbf{E}[\mathbb{1}\{X_k = j\}] = p_{jj} \xi_j. \end{aligned} \quad (11)$$

Let $\bar{Z}(K)$ denote a random variable that counts number of transitions between distinct states from time slot zero to time slot K . Note that this can be computed by subtracting all self transitions in any state from K . Therefore, we have

$$\begin{aligned} \bar{Z}(K) &= K - \sum_{j=1}^N Z_j(K) \\ \implies \lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K \bar{Z}_k}{K} &= 1 - \sum_{j=1}^N p_{jj} \xi_j. \end{aligned}$$

In the last equation, we have used (10) and (11). The LHS in the equation is equal to ν^\dagger . \square

For a two-state Markov chain, the steady-state probabilities are given by $\xi_1 = \frac{p_{21}}{p_{12} + p_{21}}$, $\xi_2 = \frac{p_{12}}{p_{12} + p_{21}}$, and $\nu^\dagger = \xi_1 p_{12} + \xi_2 p_{21} = \frac{2p_{12}p_{21}}{p_{12} + p_{21}}$. Figure 4 shows ν^\dagger versus p_{21} for different p_{12} values.

D. Unit Communication Delay

We note that the age penalty metric and the solution approach can be extended to monitoring systems where a sample can be delivered to the monitor within a unit slot time. Toward this end the age penalty at the monitor can be obtained by adding a unit delay in the difference function as given below:

$$A_k(\pi) = \max\{0, G_k - \min_n \{T_n : T_n \geq G_{k-1}\} + 1\}.$$

Given the above metric and keeping the definition of the sampling policy same as before, in the CMDP formulation for minimizing the age penalty at the monitor the per-stage cost is given by

$$c_k(X_{G_k} = j, \tau_{k+1} = \tau) = \tau + 1 - \frac{1 - p_{jj}^{\tau+1}}{1 - p_{jj}}.$$

Subsequently, with the above per-stage cost, the CMDP can be solved similarly using the Lagrangian approach. For this CMDP, the optimal sampling interval in state i is upper bounded by $\log(1 - \lambda) / \log p_{ii}$, which is one slot smaller than the upper bound given in Lemma 2.

V. NUMERICAL RESULTS

In this section, we first present numerical results for a DTMC with two-states. We then present the results for DTMCs with more than two states. The results include sampling frequency and age penalty trade-off, and a performance comparison between optimal and the best periodic sampling policies. We have implemented Algorithm 1 in MATLAB. In the following, we present two numerical examples to examine the structure of the optimal Markov policies for \mathcal{P}_1 and \mathcal{P}_2 .

Example 1: In this example, we solve \mathcal{P}_1 for transition probabilities $p_{12} = 0.1$ and $p_{21} = 0.6$, and the constraint on the expected sampling interval is equal to $1/\nu^\dagger = 5.83$. The computation of the optimal policy π_1^* results in the following stationary probabilities,

$$\begin{aligned} \mathbb{P}^{\pi_1^*}(\tau = 6|j = 1) &= 0.465 \text{ and } \mathbb{P}^{\pi_1^*}(\tau = 7|j = 1) = 0.535, \\ \mathbb{P}^{\pi_1^*}(\tau = 2|j = 2) &= 1. \end{aligned}$$

The transition probability out of state 2 is higher and thus the policy sets $\tau = 2$ when the observed state is 2. The minimum expected age penalty is computed to be 1.416. The best periodic sampling policy chooses $\tau = \tau^\dagger = \lceil 1/\nu^\dagger \rceil = 6$.

Example 2: In this example, we solve \mathcal{P}_2 by referring to the two-state DTMC in the introduction section where $p_{12} = 0.05$, $p_{21} = 0.95$, and the unit constraint for the expected age penalty implies $d = 1$. The computation of the optimal policy π_2^* results in the following stationary probabilities:

$$\begin{aligned} \mathbb{P}^{\pi_2^*}(\tau = 7|j = 1) &= 0.64 \text{ and } \mathbb{P}^{\pi_2^*}(\tau = 8|j = 1) = 0.36, \\ \mathbb{P}^{\pi_2^*}(\tau = 2|j = 2) &= 1. \end{aligned}$$

The minimum expected sampling frequency is 0.14. In contrast, using AoI as the staleness metric results in a periodic sampling policy with period 2 or sampling frequency 0.5.

A. Two-State DTMC

In Figure 5, we compare the average age penalties achieved by the best periodic sampler and the optimal policy π_1^* obtained by solving \mathcal{P}_1 under the constraint $\nu = \nu^\dagger$. Recall that for this case, the best periodic sampler sets the sampling interval equal to $\tau^\dagger = \lceil 1/\nu^\dagger \rceil$. From the figure, we observe that for lower transition probabilities between the states, i.e., lower p_{12} and p_{21} values, the periodic sampler achieves age penalties only slightly higher than that of the optimal policy, because in this case the optimal policy is also choosing sampling intervals close to that of the periodic sampler. The gap between them, however, increases significantly for higher transition probabilities. The zigzag pattern of the periodic sampler can be attributed to the ceiling function used in computing the sampling interval.

In Figures 6, and 7 we compare average sampling frequencies achieved by the best periodic sampler and the optimal

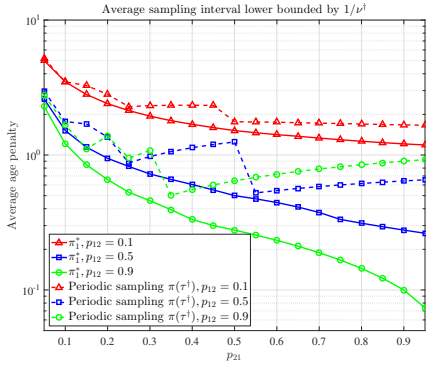


Fig. 5: Average age penalties achieved by π_1^* and the best periodic sampler for different p_{12} and p_{21} values.

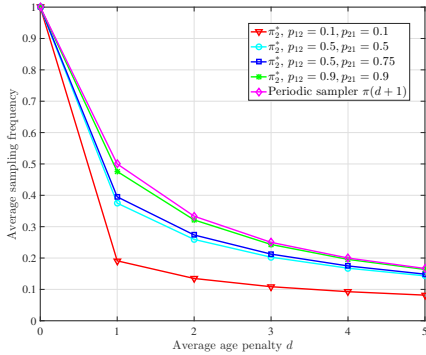


Fig. 6: Sampling frequency and age penalty trade-off: average sampling frequencies achieved by π_2^* and the best periodic sampler by varying age penalty constraint value.

policy π_2^* by solving \mathcal{P}_2 . From Figure 6, we observe the trade-off between achievable sampling frequencies and age penalties. As expected, for an age penalty constraint of one time slot, i.e. $d = 1$, the achievable sampling frequency is lower than 0.5 for both policies. However, π_2^* results in much lower sampling frequencies for lower transition probabilities. In Figure 7, we set $d = 1$ and thus the best periodic sampler samples every 2 time slots with sampling frequency 0.5. On the other hand, π_2^* provides much lower sampling frequencies when either of the transition probabilities are small.

Finally, in Figure 8, we present the ratio between the expected sampling frequency achieved by π_2^* and ν^\dagger , under average age penalty constraint $d = 1$. We note that under π_1^\dagger the age penalty is always zero. This cannot be achieved by any causal policy with a sampling frequency strictly less than one. Nonetheless, an interesting observation from the figure is that by allowing a small age penalty $d = 1$, the optimal policy π_2^* can achieve lower sampling frequency than ν^\dagger when transition probabilities are higher, say $p_{12} = 0.9$ and $p_{21} = 0.9$. For lower transition probabilities $p_{12} = 0.1$ and $p_{21} = 0.1$, the ratio is always greater than 1, i.e., the optimal policy π_2^* could not achieve the sampling frequency ν^\dagger and may require more relaxation in the age penalty constraint. In conclusion, for lower transition probabilities, i.e., if the events become rare, the optimal policy performs worse with respect to the optimal clairvoyant sampling policy.

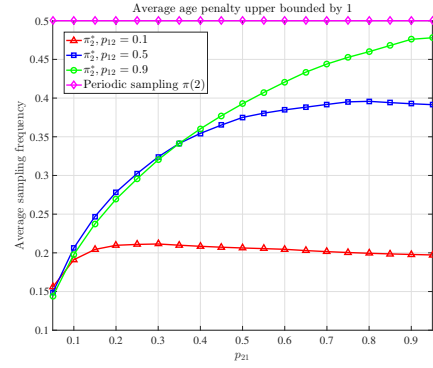


Fig. 7: Average sampling frequencies achieved by π_2^* and the best periodic sampler for varying p_{12} , and $d = 1$.

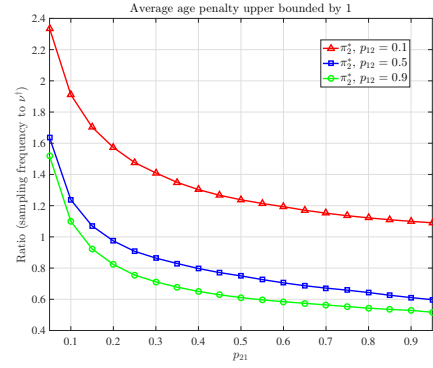


Fig. 8: Ratio between the expected sampling frequency achieved by π_2^* and ν^\dagger , and $d = 1$.

B. N-State DTMC

For the N -state DTMC, with $N > 2$, it would be a tedious task to present the results by varying different transition probabilities independently. Instead, we choose a transition probability matrix, parameterized by N and a probability p , with p_{ij} defined as:

$$p_{ij} = \begin{cases} p & \text{if } i = j \\ \frac{1-p}{N-1} & \text{otherwise.} \end{cases}$$

Note that, for the two-state DTMC, we have $p_{12} = 1 - p$ and $p_{21} = 1 - \frac{p}{2}$. Also, by increasing p , we decrease the probability of transitions to the other states, which implies ν^\dagger increases. However, the relation between ν^\dagger and N is not easy to infer from the above transition probabilities, because the transition probabilities to other states decrease as N increases, but also the self-state transition probabilities of states with larger index decreases. However, from Figure 9 we observe that ν^\dagger increases as N increases.

In Figure 10, we present the average age penalty achieved by π_1^* and the best periodic policy by varying p . We observe that the optimal age penalty decreases as N increases. This is because ν^\dagger increases as N increases, and since $\frac{1}{\nu^\dagger}$ is the lower-bound constraint on the average sampling interval, the optimal age penalty decreases. Similarly, as p increases the optimal age penalty increases as ν^\dagger decreases. Also, observe that the optimal policy π_1^* provides a lower age penalty by at least half that of the age penalty under the best periodic sampling policy.

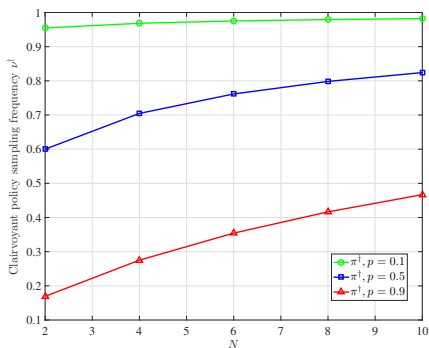


Fig. 9: ν^\dagger achieved by the clairvoyant optimal policy π^\dagger for varying N and for different p values.

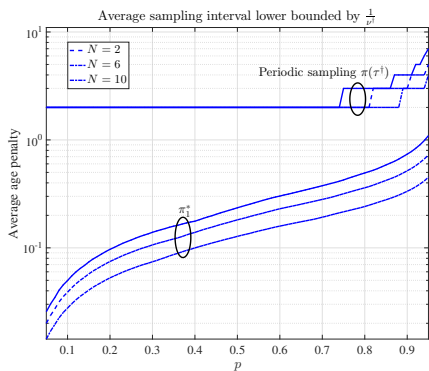


Fig. 10: Avg. age penalty versus p for different values of N .

Similar reduction in average sampling frequency achieved by π_2^* is observed from Figure 11, where we set the age limit equal to 1.

In Figure 12, we present the trade-off between the minimum sampling frequency and the age penalty for $N = 6$. We observe similar trends as in the case of $N = 2$ in Figure 6 where π_2^* results in a reduction of more than 70% by allowing an age limit 1 for higher values for p , i.e., lower transition probabilities to other states. Finally, in Figure 13, we present the ratio between the minimum sampling frequency achieved by π_2^* and ν^\dagger by varying N . Note that, by allowing age limit 1, we not only achieve lower sampling frequency than ν^\dagger , but also the reduction factor in the ratio increases as N increases. We conclude that, for the given parameterized transition probabilities, the results for $N > 2$ follow similar trends as in the case of $N = 2$, and also in several cases the reductions are higher for larger N .

VI. CONCLUSION

We have studied the trade-off between sampling frequency and staleness for detecting transitions of a DTMC to new states. The staleness of the k th sample is quantified using age penalty, which is defined as the time elapsed since the first transition out of the state in the $k - 1$ sample. The formulated problems \mathcal{P}_1 and \mathcal{P}_2 are CMDPs and were solved by a method where we randomize between two deterministic policies, which are obtained by solving the Lagrangian MDPs using RVI. The proven structural property of the solution has

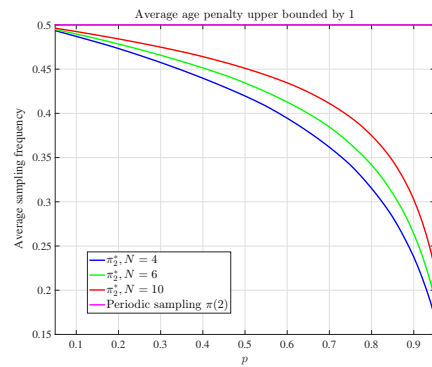


Fig. 11: Avg. sampling frequency achieved by π_2^* and the best periodic sampler for varying p and for different values of N .

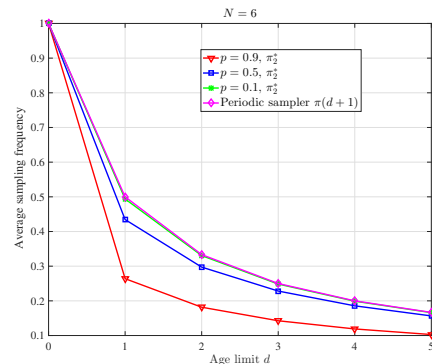


Fig. 12: Sampling frequency and age penalty trade-off: average sampling frequencies achieved by π_2^* and the best periodic sampler by varying the age limit d .

been exploited to reduce the computational complexity of the proposed method. We have provided a closed-form expression for ν^\dagger , the sampling frequency under the optimal clairvoyant sampling policy. Apart from the superior performance of the computed optimal policy over the best periodic sampling policy, we found that by allowing a small age penalty the optimal policy achieves a sampling frequency lower than ν^\dagger in some scenarios. In future work, we plan to study the trade-off by considering age-penalty at the remote monitor and under different communication models.

REFERENCES

- [1] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Towards an effective age of information: Remote estimation of a markov source," in *IEEE INFOCOM WKSHPs*, April 2018, pp. 367–372.
- [2] C. Kam, S. Kompella, and A. Ephremides, "Age of incorrect information for remote estimation of a binary markov source," in *Proc. IEEE INFOCOM WKSHPs*, 2020, pp. 1–6.
- [3] Y. Chen and A. Ephremides, "Minimizing age of incorrect information for unreliable channel with power constraint," *CoRR*, vol. abs/2101.08908, 2021.
- [4] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, p. 2215–2228, Oct. 2020.
- [5] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. IEEE SECON*, 2011.
- [6] V. Krishnamurthy, *Partially Observed Markov Decision Processes*. Cambridge University Press, 2016.
- [7] E. Altman, *Constrained Markov Decision Processes*. Chapman and Hall, 1999.

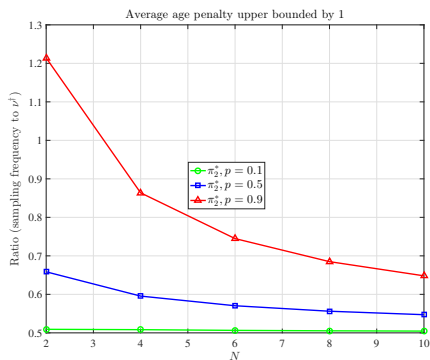


Fig. 13: Ratio between the expected sampling frequency achieved by π_2^* and ν^\dagger .

- [8] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," *CoRR*, vol. abs/1806.06243, 2018.
- [9] S. Feng and J.-S. Yang, "Information freshness for timely detection of status changes," *ArXiv*, vol. abs/2002.04648, 2020.
- [10] Y. Inoue and T. Takine, "AoI perspective on the accuracy of monitoring systems for continuous-time markovian sources," in *IEEE INFOCOM WKSHPs*, April 2019, pp. 183–188.
- [11] S. Kriouile and M. Assaad, "Minimizing the age of incorrect information for real-time tracking of markov remote sources," *CoRR*, vol. abs/2102.03245, 2021. [Online]. Available: <https://arxiv.org/abs/2102.03245>
- [12] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, Nov 2017.
- [13] E. T. Ceran, D. Gündüz, and A. Gyöngy, "Average age of information with hybrid arq under a resource constraint," in *Proc. WCNC*, 2018, pp. 1–6.
- [14] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," in *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, 2019, pp. 1–8.
- [15] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the internet of things," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7468–7482, 2019.
- [16] H. Tang, J. Wang, L. Song, and J. Song, "Minimizing age of information with power constraints: Multi-user opportunistic scheduling in multi-state time-varying channels," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 854–868, 2020.
- [17] H. Kushner, "On the optimum timing of observations for linear control systems with unknown initial state," *IEEE Transactions on Automatic Control*, vol. 9, no. 2, pp. 144–150, April 1964.
- [18] E. Skafidas and A. Nerode, "Optimal measurement scheduling in linear quadratic gaussian control problems," in *Proceedings of the 1998 IEEE International Conference on Control Applications (Cat. No.98CH36104)*, vol. 2, Sep. 1998, pp. 1225–1229 vol.2.
- [19] M. Rabi, G. V. Moustakides, and J. S. Baras, "Adaptive sampling for linear state estimation," *SIAM Journal on Control and Optimization*, vol. 50, no. 2, pp. 672–702, 2012.
- [20] R. P. Anderson, D. Milutinović, and D. V. Dimarogonas, "Self-triggered sampling for second-moment stability of state-feedback controlled sde systems," *Automatica*, vol. 54, pp. 8 – 15, 2015.
- [21] J. Cho and H. Garcia-Molina, "Effective page refresh policies for web crawlers," *ACM Trans. Database Syst.*, vol. 28, no. 4, p. 390–426, Dec. 2003.
- [22] C. Olston and M. Najork, "Web crawling," *Found. Trends Inf. Retr.*, vol. 4, no. 3, p. 175–246, Mar. 2010.
- [23] K. Sigman, Markov chains II: recurrence and limiting (stationary) distributions. Lecture Notes on Stochastic Modeling I. [Online]. Available: <http://www.columbia.edu/~ks20/stochastic-I/stochastic-I.html>
- [24] A. S. Manne, "Linear programming and sequential decisions," *Management Science*, vol. 6, no. 3, pp. 259–267, 1960.
- [25] F. J. Beutler and K. W. Ross, "Optimal policies for controlled markov chains with a constraint," *Journal of Mathematical Analysis and Applications*, vol. 112, no. 1, pp. 236–252, 1985.
- [26] D. Ma, A. M. Makowski, and A. Shwartz, "Estimation and optimal

control for constrained markov chains," in *Proc. IEEE CDC*, 1986, pp. 994–999.

- [27] J. P. Champati, M. Skoglund, and J. Gross, "Detecting state transitions of a markov source: Sampling frequency and age trade-off," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs)*, 2020, pp. 7–12.
- [28] J. van den Brand, "A deterministic linear program solver in current matrix multiplication time," in *Proc. ACM SODA*. Society for Industrial and Applied Mathematics, 2020, p. 259–278.
- [29] S. Boyd and L. Vandenberghe, *Convex Optimization*. USA: Cambridge University Press, 2004.
- [30] J. P. Champati, H. Al-Zubaidy, and J. Gross, "On the distribution of aoi for the gi/gi/1/1 and gi/gi/1/2* systems: Exact expressions and bounds," in *Proc. IEEE INFOCOM*, 2019, pp. 37–45.
- [31] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol. I*, 4th ed. Athena Scientific, 2017.
- [32] R. Gallager, *Discrete Stochastic Processes*, ser. The Springer International Series in Engineering and Computer Science. Springer US, 2012.

APPENDIX

A. Proof of Lemma 2

1) Binary-Decision CMDP Formulation:

- **State space:** In time slot k , the state of the binary decision CMDP is given by $Y_k = (\bar{X}_k, Z_k)$, where $\bar{X}_k \in S = \{1, 2, \dots, N\}$ is the most recent state of the DTMC that was observed and $Z_k \in Q = \{1, 2, \dots, M\}$ is the number of slots elapsed since \bar{X}_k was observed until the current slot k . The state space for Y_k is given by $S \times Q$.
- **Action space:** In time slot k , a binary action $u_k \in \{0, 1\}$ is taken, where $u_k = 0$ represents no-sampling, and $u_k = 1$ represents sampling in slot k .
- **Transition probabilities:** In time slot $k+1$, the next state $Y_{k+1} = (\bar{X}_{k+1}, Z_{k+1})$ only depends on the current state $Y_k = (\bar{X}_k, Z_k)$ and action u_k . If $Y_k = (i, z)$ and $u_k = 0$, then $\bar{X}_{k+1} = i$, since no new sample is taken, and $Z_{k+1} = z + 1$. Therefore, in this case, we have $Y_{k+1} = (i, z+1)$. On the other hand, if $u_k = 1$, then $\bar{X}_{k+1} = X_k$, the state of the DTMC observed from the sample in slot k , and $Z_{k+1} = 1$. As a result, we arrive at the following transition probabilities:

$$\mathbb{P}(Y_{k+1}=(j, \bar{z}) \mid Y_k=(i, z), u_k=0) = \begin{cases} 1 & \text{if } j=i, \bar{z}=z+1, \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbb{P}(Y_{k+1}=(j, \bar{z}) \mid Y_k=(i, z), u_k=1) = \begin{cases} p_{ij}^{(z)} & \text{if } \bar{z}=1, \\ 0 & \text{otherwise.} \end{cases}$$

- **Costs:** In slot k , if $Y_k = (i, z)$, then the per-stage cost, denoted by $\gamma_k((i, z), u_k)$, contributing to the average age-penalty is given by

$$\gamma_k((i, z), u_k) = (1 - u_k)(1 - p_{ii}^{z-1}).$$

and the cost contributing to the average sampling-interval is given by u_k .

For $u \in \{0, 1\}$, a Markov policy π_b specifies the conditional distribution $\mathbb{P}^{\pi_b}(u \mid (i, z))$ for all states (i, z) . Note that we use the subscript b to differentiate the policy of binary-decision CMDP from that of the policy π defined in Section III that maps states to sampling intervals. Let Π_b^{MR} denote the set of

Markov policies constituting all feasible π_b . We now state the following problem:

$$\begin{aligned} & \underset{\pi_b \in \Pi_b^{\text{MR}}}{\text{minimize}} && \lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}^{\pi_b} \left[\sum_{k=1}^K \gamma_k(Y_k, u_k) \right] \\ & \text{s.t.} && \lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}^{\pi_b} \left[\sum_{k=1}^K u_k \right] \leq \nu, \end{aligned} \quad (12)$$

where the expectations above are taken with respect to the induced probability distribution under π_b . We claim that \mathcal{P}_1 and (12) are equivalent. To prove this, we proceed as follows. Any policy $\hat{\pi}_b \in \Pi_b^{\text{MR}}$ can be converted to an *equivalent policy* $\hat{\pi} \in \Pi^{\text{MR}}$ and vice-versa. Given $\hat{\pi}_b$, $\hat{\pi}$ is constructed as follows. For all (i, z) ,

$$\mathbb{P}^{\hat{\pi}}(\tau = z|i) = \frac{\mathbb{P}^{\hat{\pi}_b}(u = 1|(i, z))}{\sum_{\bar{z}=1}^M \mathbb{P}^{\hat{\pi}_b}(u = 1|(i, \bar{z}))}. \quad (13)$$

However, for constructing a $\hat{\pi}_b$ for a given π , we do not need a normalization constant used in the denominator of the above equation.

$$\mathbb{P}^{\hat{\pi}_b}(u = 1|(i, z)) = \mathbb{P}^{\hat{\pi}}(\tau = z|i) \quad (14)$$

From (13) and (14) any feasible solution to \mathcal{P}_1 results in a feasible solution to (12) and vice-versa.

Given state i is observed in the current slot, $\hat{\pi}$ chooses a sampling interval equal to z (with probability $\mathbb{P}^{\hat{\pi}}(\tau = z|i)$), then it incurs an age penalty equal to $z - \frac{1-p_{ii}^z}{1-p_{ii}}$ (cf. (4)) and the sampling cost is z . Similarly, under $\hat{\pi}_b$, if state i is observed in the current slot, and a sample is taken in z -th slot from the current time slot (with probability $\mathbb{P}^{\hat{\pi}_b}(u = 1|(i, z))$), it also incurs a sampling cost z and cumulative penalty equals $\sum_{k=1}^z (1 - p_{ii}^{k-1})$, which is equal to $z - \frac{1-p_{ii}^z}{1-p_{ii}}$. Thus, the expected age penalty and the expected frequency achieved under $\hat{\pi}_b$ and $\hat{\pi}$ are equal. Therefore, \mathcal{P}_1 and (12) are equivalent. Given this, we prove the result for the Lagrangian MDP for the binary-decision CMDP, which is given by

$$\underset{\pi_b \in \Pi_b^{\text{MR}}, \lambda \geq 0}{\text{minimize}} \quad \lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}^{\pi_b} \left[\sum_{k=1}^K \{\gamma_k(Y_k, u_k) + \lambda(u_k - \nu)\} \right]. \quad (15)$$

The per-stage cost of the Lagrangian MDP, denoted by $\gamma_k^\lambda((i, z), u_k)$, is given by

$$\gamma_k^\lambda((i, z), u_k) = (1 - u_k)(1 - p_{ii}^{z-1}) + u_k \lambda. \quad (16)$$

Note that $1 - p_{ii}^{z-1} < 1$, for all i and $z < \infty$. Therefore, if $\lambda \geq 1$, then from (16), we infer that the optimal decision is to set $u_k = 0$, and this is true for any k . Since $\lambda \geq 1$ results in a trivial policy, it is sufficient to restrict the search space for λ in (15) to $0 < \lambda < 1$.

Given $\lambda \geq 0$, there exists an optimal deterministic policy for the Lagrangian MDP. Further, it is known for a finite-state finite-action space CMDP with a single constraint, there exists two distinct values for λ , say λ_1 and λ_2 , such that an optimal Markov policy for binary-decision CMDP can be obtained by randomizing between the optimal deterministic policies for the two Lagrangian MDPs corresponding to λ_1 and λ_2 .

Lemma 4. Let z_i^* denote the optimal inter-sampling time in state $(i, 1)$ for the Lagrangian MDP with parameter λ , then we have $z_i^* < \log(1 - \lambda) / \log p_{ii} + 1$.

We use Value Iteration (VI) algorithm [31] to prove that $z_i^* < \log(1 - \lambda) / \log p_{ii} + 1$ or $1 - p_{ii}^{z_i^*-1} > \lambda$. Let $J_l(i, z)$ denote the l -stage cost in the VI algorithm for state (i, z) . If the decision in $(l + 1)$ -th-stage is $u_l = 0$, then the new state will be $(i, z + 1)$, and the l -th stage cost will be $J_l(i, z + 1)$. If $u_l = 1$, then the new state $(i', 1)$, for some $i' \in Q$, and the l -th stage cost will be $\sum_{i' \in Q} p_{ii'}^{(\tau)} J_k(i', 1)$. Therefore, we have

$$\begin{aligned} J_{l+1}(i, z) = & \min_{u \in \{0,1\}} \left[\gamma_l^\lambda((i, z), u) + (1 - u)J_l(i, z + 1) \right. \\ & \left. + u \sum_{i' \in Q} p_{ii'}^{(z)} J_k(i', 1) \right], \end{aligned} \quad (17)$$

where $\gamma_l^\lambda(\cdot)$ is given in (16). Let $u_l^*(i, z)$ denote the optimum solution to (17).

Since VI algorithm converges to the optimal solution starting with any terminal cost $J_0(i, z)$, we set $J_0(i, z) = 0$ for all (i, z) . Now, for all z such that $1 - p_{ii}^{z-1} > \lambda$, we show that the VI algorithm results in $u_l^*(i, z) = 1$, for all $l \geq 0$. For $l = 0$, we have

$$\begin{aligned} J_1(i, z) &= \min_{u \in \{0,1\}} [\gamma_0^\lambda((i, z), u)] \\ &= \min_{u \in \{0,1\}} (1 - u)(1 - p_{ii}^{z-1}) + u\lambda = \lambda. \end{aligned}$$

The last step above is true for any z such that $1 - p_{ii}^{z-1} > \lambda$, and we have $u_0^*(i, z) = 1$. For $l > 0$, we use proof by induction. Assume that, for all z such that $1 - p_{ii}^{z-1} > \lambda$, $u_l^*(i, z) = 1$ is true for some $l > 0$ and, then we have

$$J_l(i, z) = \gamma_l^\lambda((i, z), 1) + \sum_{i' \in Q} p_{ii'}^{(z)} J_{k-1}(i', 1) \quad (18)$$

If $u_{l+1}^*(i, z) = 1$ were to be true, then from (17) it should be true that

$$\gamma_l^\lambda((i, z), 1) + \sum_{i' \in Q} p_{ii'}^{(z)} J_l(i', 1) \leq \gamma_l^\lambda((i, z), 0) + J_l(i, z + 1). \quad (19)$$

Note that (18) is valid for $z + 1$ since $1 - p_{ii}^z > \lambda$. Therefore, we have

$$\begin{aligned} J_l(i, z + 1) &= \gamma_l^\lambda((i, z + 1), 1) + \sum_{i' \in Q} p_{ii'}^{(z+1)} J_{l-1}(i', 1) \\ &= \lambda + \sum_{i' \in Q} p_{ii'}^{(z+1)} J_{l-1}(i', 1). \end{aligned} \quad (20)$$

Furthermore, using $u = 1$ for $J_l(i', 1)$, we obtain

$$J_l(i', 1) \leq \lambda + \sum_j p_{i'j}^{(z)} J_{l-1}(j, 1). \quad (21)$$

Substituting $\gamma_l^\lambda((i, z), 1) = \lambda$ and (21) in the LHS of (19), we obtain

$$\gamma_l^\lambda((i, z), 1) + \sum_{i' \in Q} p_{ii'}^{(z)} J_l(i', 1) \leq 2\lambda + \sum_j p_{ij}^{(z+1)} J_{l-1}(j, 1) \quad (22)$$

Substituting $\gamma_i^\lambda((i, z), 0) = 1 - p_{ii}^{z-1}$ and (20) in the RHS of (19), we obtain

$$\begin{aligned} & \gamma_i^\lambda((i, z), 0) + J_l(i, z + 1) \\ &= 1 - p_{ii}^{z-1} + \lambda + \sum_{i' \in Q} p_{ii'}^{(z+1)} J_{l-1}(i', 1) \end{aligned} \quad (23)$$

The RHS in (23) is greater than the RHS in (22). Therefore, the inequality in (19) is true.

B. Proof of Lemma 3

For this proof, we formulate the CMDP as a linear program (cf. [7]) and show that the constraint in the linear program is tight.

LP formulation of \mathcal{P}_1 : We define $c_{j\tau} = \sum_{n=1}^{\tau-1} (\tau - n)(1 - p_{jj})p_{jj}^{n-1}$. Further, we define $z_{j\tau}^\pi = \mathbb{P}(X_{G_k} = j, \tau_k = \tau)$, the steady-state probability of observing the state-action pair (j, τ) under a policy $\pi \in \Pi^{\text{MR}}$. Then, we obtain

$$\begin{aligned} \mathbb{E}[A(\pi)] &= \sum_{j=1}^N \sum_{\tau=1}^M c_{j\tau} z_{j\tau}^\pi \\ \limsup_{K \rightarrow \infty} \frac{\mathbb{E}[\sum_{k=1}^K \tau_k]}{K} &= \sum_{j=1}^N \sum_{\tau=1}^M \tau z_{j\tau}^\pi. \end{aligned}$$

In the LP formulations of \mathcal{P}_1 and \mathcal{P}_2 , we solve for $z_{j\tau}^\pi$ with the following constraints,

$$\sum_{j=1}^N \sum_{\tau=1}^M z_{j\tau}^\pi = 1, \quad (24)$$

$$\sum_{\tau=1}^M z_{i\tau}^\pi = \sum_{j=1}^N \sum_{\tau=1}^M q_{j\tau i} z_{j\tau}^\pi, \quad i \in S, \quad (25)$$

$$z_{j\tau}^\pi \geq 0, \quad j \in S \text{ and } \tau \in A. \quad (26)$$

The constraint (25) is a consequence of the equilibrium equations for the induced DTMC in the steady state. In the following, we present an equivalent LP formulation of \mathcal{P}_1 ,

$$\begin{aligned} & \underset{\{z_{j\tau}^\pi\}}{\text{minimize}} \quad \sum_{j=1}^N \sum_{\tau=1}^M c_{j\tau} z_{j\tau}^\pi \\ & \text{s.t.} \quad \sum_{j=1}^N \sum_{\tau=1}^M \tau z_{j\tau}^\pi \geq \frac{1}{\nu}, \quad (27) \\ & \quad \quad \quad (24), (25), (26). \end{aligned}$$

Let \mathbf{z}^* denote the optimal solution for (27), then the stationary probabilities under π_1^* are computed as follows. For $\tau \in A$,

$$\mathbb{P}^{\pi_1^*}(\tau|j) = \frac{z_{j\tau}^*}{\sum_{\tau=1}^M z_{j\tau}^*}, \quad j \in S.$$

We prove the lemma by proving that

$$\sum_{j=1}^N \sum_{\tau=1}^M \tau z_{j\tau}^* = \frac{1}{\nu}. \quad (28)$$

Towards this end, we use proof by contradiction. Assume that (28) is not true and for some $0 < \nu^* < \nu$ we have

$$\sum_{j=1}^N \sum_{\tau=1}^M \tau z_{j\tau}^* = \frac{1}{\nu^*} > \frac{1}{\nu}.$$

We construct another solution $\bar{\mathbf{z}} = \{\bar{z}_{j\tau}\}$ using \mathbf{z}^* such that it has lower expected age penalty and lower expected inter-sampling time. Let $\hat{\tau}$ denote the maximum value of τ for which $z_{j\tau}^*$ is non-zero for some j , i.e., $\hat{\tau} = \max\{\tau : \exists j \text{ such that } z_{j\tau}^* > 0\}$. The key to our proof is the following construction. Recall that, ζ_j is the steady-state probability of finding the DTMC in state j . Let $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_N]^\top$ and $\boldsymbol{\epsilon}' = [\epsilon'_1, \epsilon'_2, \dots, \epsilon'_N]^\top$ denote two non-negative column vectors, then $\bar{z}_{j\tau}$ are constructed as follows:

$$\begin{aligned} \bar{z}_{j\tau} &:= z_{j\tau}^*, \quad \forall j, \tau \notin \{\hat{\tau} - 1, \hat{\tau}\} \\ \bar{z}_{j\hat{\tau}} &:= z_{j\hat{\tau}}^* - \xi_j \epsilon_j, \quad \forall j \\ \bar{z}_{j\hat{\tau}-1} &:= z_{j\hat{\tau}-1}^* + \xi_j \epsilon'_j, \quad \forall j. \end{aligned} \quad (29)$$

We next focus on the set of conditions $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}'$ should satisfy such that the constraints (24), (25), and (26) are satisfied by $\bar{\mathbf{z}}$.

Satisfying Constraint (24): Since \mathbf{z}^* is a solution to (27), it satisfies (24). Using this, we have

$$\begin{aligned} \sum_{j=1}^N \sum_{\tau=1}^M \bar{z}_{j\tau} &= \sum_{j=1}^N \sum_{\tau=1}^M z_{j\tau}^* - \sum_{j=1}^N \xi_j \epsilon_j + \sum_{j=1}^N \xi_j \epsilon'_j \\ &= 1 - \sum_{j=1}^N \xi_j \epsilon_j + \sum_{j=1}^N \xi_j \epsilon'_j. \end{aligned}$$

Therefore, for $\bar{\mathbf{z}}$ to satisfy (24), we should have

$$\boldsymbol{\xi} \boldsymbol{\epsilon}' = \boldsymbol{\xi} \boldsymbol{\epsilon}. \quad (30)$$

Satisfying Constraint (26): Since $z_{j\tau}^* \geq 0$, for all j and for all τ , it is sufficient to choose $\boldsymbol{\epsilon}$ appropriately such that $\bar{z}_{j\hat{\tau}} \geq 0$, for all j . To this end, we describe a procedure in Algorithm 2. We define η , used in line 4 of Algorithm 2, as follows.

$$\eta = \min \left\{ \frac{1}{\nu^*} - \frac{1}{\nu}, \sum_{j=1}^N z_{j\hat{\tau}}^* \right\}.$$

Since $\nu^* < \nu$, and $z_{j\hat{\tau}}^* > 0$ for some j , we have $\eta > 0$. In Algorithm 2, we describe a procedure for assigning $\boldsymbol{\epsilon}$. It

Algorithm 2: Procedure for choosing $\boldsymbol{\epsilon}$

- 1: If $\frac{1}{\nu^*} - \frac{1}{\nu} > \sum_{j=1}^N z_{j\hat{\tau}}^*$ then $\epsilon_j = \frac{z_{j\hat{\tau}}^*}{\xi_j}$ for all j , otherwise do the following.
 - 2: $h \leftarrow \eta$
 - 3: **for** $j \leftarrow 1$ **to** N **do**
 - 4: **if** $z_{j\hat{\tau}}^* = 0$, **then** $\epsilon_j = 0$, otherwise do the following.
 - 5: **if** $z_{j\hat{\tau}}^* \leq h$, **then** $\epsilon_j = \frac{z_{j\hat{\tau}}^*}{\xi_j}$ and $h \leftarrow h - z_{j\hat{\tau}}^*$, **else**
 $\epsilon_j = z_{j\hat{\tau}}^* - h$.
 - 6: **end for**
-

is easy to verify that Algorithm 2 assigns values to ϵ_j such that $\epsilon_j \leq z_{j\hat{\tau}}^*$, for all j . Using this in (29), we infer that $\bar{z}_{j\hat{\tau}} \geq 0$, for all j and for all τ . Also, it can be verified that $\boldsymbol{\xi} \boldsymbol{\epsilon} = \eta$. Therefore, from (30), we should have $\boldsymbol{\xi} \boldsymbol{\epsilon}' = \eta$. By the definition of $\hat{\tau}$, we have $z_{j\tau}^* = 0$ for all $\tau > \hat{\tau}$ and therefore, $\bar{z}_{j\tau} = 0$ for all $\tau > \hat{\tau}$. Using this, we have

$$\sum_{j=1}^N \sum_{\tau=1}^M \tau \bar{z}_{j\tau} = \sum_{j=1}^N \sum_{\tau=1}^{\hat{\tau}} \tau \bar{z}_{j\tau}$$

$$\begin{aligned}
&= \sum_{j=1}^N \sum_{\tau=1}^{\hat{\tau}-2} \tau z_{j\tau}^* + (\hat{\tau}-1) \sum_{j=1}^N (z_{j\hat{\tau}-1}^* + \xi_j \epsilon_j') + \hat{\tau} \sum_{j=1}^N (z_{j\hat{\tau}}^* - \xi_j \epsilon_j) \\
&= \sum_{j=1}^N \sum_{\tau=1}^{\hat{\tau}} \tau z_{j\tau}^* + (\hat{\tau}-1)\eta - \hat{\tau}\eta = \frac{1}{\nu^*} - \eta > \frac{1}{\nu}.
\end{aligned}$$

Therefore, $\bar{\mathbf{z}}$ has strictly lower expected inter-sampling time than that of \mathbf{z}^* .

Satisfying Constraint (25): Given that ϵ is obtained using Algorithm 2, we now obtain a set of equations to be satisfied by ϵ' , so that $\{\bar{z}_{i\tau}\}$ satisfies (25). For all $i \in S$,

$$\begin{aligned}
&\sum_{\tau=1}^M \bar{z}_{i\tau} = \sum_{j=1}^N \sum_{\tau=1}^M p_{ji}^{(\tau)} \bar{z}_{j\tau} \\
\Rightarrow &\sum_{\tau=1}^{\hat{\tau}} z_{i\tau}^* + \xi_i \epsilon_i' - \xi_i \epsilon_i \\
&= \sum_{j=1}^N \sum_{\tau=1}^{\hat{\tau}} p_{ji}^{(\tau)} z_{j\tau}^* + \sum_{j=1}^N p_{ji}^{(\hat{\tau}-1)} \xi_j \epsilon_j' - \sum_{j=1}^N p_{ji}^{(\hat{\tau})} \xi_j \epsilon_j \\
\Rightarrow &\sum_{j=1, j \neq i}^N p_{ji}^{(\hat{\tau}-1)} \xi_j \epsilon_j' + (p_{ii}^{(\hat{\tau}-1)} - 1) \xi_i \epsilon_i' \\
&= \sum_{j=1, j \neq i}^N p_{ji}^{(\hat{\tau})} \xi_j \epsilon_j + (p_{ii}^{(\hat{\tau})} - 1) \xi_i \epsilon_i, \quad \forall i \in S. \quad (31)
\end{aligned}$$

Unique solution: We show that there exists a unique solution for the systems of $N+1$ linear equations which constitute (30) and (31). The system matrix $\mathbf{H} = \{h_{ij}\}$ is of dimension $(N+1) \times N$, and its ij -th element h_{ij} is given by

$$\begin{aligned}
h_{ij} &= p_{ji}^{(\hat{\tau}-1)} \xi_j, & i \neq j, j \leq N, i \leq N \\
h_{ii} &= (p_{ii}^{(\hat{\tau}-1)} - 1) \xi_i, & i \leq N \\
h_{(N+1)j} &= \xi_j, & j \leq N.
\end{aligned}$$

Let \mathbf{r} denote a vector of dimension $N+1 \times 1$ with i -th element r_i given by

$$r_i = \sum_{j=1, j \neq i}^N p_{ji}^{(\hat{\tau})} \xi_j \epsilon_j + (p_{ii}^{(\hat{\tau})} - 1) \xi_i \epsilon_i, \quad \forall i \leq N, \quad r_{N+1} = \xi \epsilon.$$

The system of linear equations can then be written in the following form: $\mathbf{H}\epsilon' = \mathbf{r}$.

Recall that $p_{ji}^{(\hat{\tau}-1)}$ is the ji -th element of $P^{(\hat{\tau}-1)}$. Let $\mathbf{Q}_{\hat{\tau}} = (P^{(\hat{\tau}-1)})^\top$, $\mathbf{D}_\xi = \text{diag}(\xi)$, and \mathbf{I} denote identity matrix, then \mathbf{H} can be written as

$$\mathbf{H} = \begin{bmatrix} \mathbf{D}_\xi(\mathbf{Q}_{\hat{\tau}} - \mathbf{I}) \\ \xi \end{bmatrix}.$$

Since P is the transition probability matrix of an ergodic DTMC, $\mathbf{Q}_{\hat{\tau}}^\top$ will also be a transition probability matrix of an ergodic DTMC. To see this, a DTMC is ergodic if and only if its transition probability matrix P when multiplied by itself repeatedly converges to a positive matrix with identical rows [32]. Since P^n converges to a matrix with rows equal to ξ as n goes to infinity, so does $(\mathbf{Q}_{\hat{\tau}}^\top)^n$. This further implies

that, ξ is a unique solution that satisfies $\xi = \xi \mathbf{Q}_{\hat{\tau}}^\top$, which is stated in the following lemma.

Lemma 5. ξ^\top is the unique normalized solution \mathbf{x} to the system of equations $(\mathbf{Q}_{\hat{\tau}} - \mathbf{I})\mathbf{x} = 0$.

In the following lemma, we present a result on the rank of the matrix \mathbf{H} .

Lemma 6. The matrix \mathbf{H} has rank N .

Proof. Let $\text{Null}(\mathbf{B})$ denote the null space of matrix \mathbf{B} . We first prove $\text{Null}(\mathbf{D}_\xi(\mathbf{Q}_{\hat{\tau}} - \mathbf{I}))$ has only one basis vector. To see this, consider the solution set of

$$\mathbf{D}_\xi(\mathbf{Q}_{\hat{\tau}} - \mathbf{I})\mathbf{x} = 0 \iff (\mathbf{Q}_{\hat{\tau}} - \mathbf{I})\mathbf{x} = 0.$$

In the second step above, we have used the fact that the diagonal elements of \mathbf{D}_ξ are strictly positive and it is invertible. Therefore, from Lemma 5 we conclude that ξ^\top is the unique non-zero solution for $\mathbf{D}_\xi(\mathbf{Q}_{\hat{\tau}} - \mathbf{I})\mathbf{x} = 0$. This implies that ξ^\top is the only basis for $\text{Null}(\mathbf{D}_\xi(\mathbf{Q}_{\hat{\tau}} - \mathbf{I}))$. Since the row space and the null space of $\mathbf{D}_\xi(\mathbf{Q}_{\hat{\tau}} - \mathbf{I})$ spans \mathbb{R}^N , the N dimensional real space, the row vectors of $\mathbf{D}_\xi(\mathbf{Q}_{\hat{\tau}} - \mathbf{I})$ and ξ form a basis for \mathbb{R}^N . In other words, \mathbf{H} , whose rows consist of ξ and the rows of $\mathbf{D}_\xi(\mathbf{Q}_{\hat{\tau}} - \mathbf{I})$, has row space equal to \mathbb{R}^N ; thus, its rank is N . \square

From Lemma 6, we infer that the columns of \mathbf{H} are independent which implies that $\mathbf{H}^\top \mathbf{H}$ is invertible. Therefore, we obtain a unique solution $\epsilon' = (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{r}$. This implies that $\bar{\mathbf{z}}$ can be constructed such that it is a feasible solution to the CMDP and also has lower expected age penalty than \mathbf{z}^* , which proves the contradiction.