Timely Monitoring of Markov Chains Under Sampling Rate Constraints

Nail Akar
Electrical and Electronics Engineering Dept.
Bilkent University
Ankara, Turkey
akar@ee.bilkent.edu.tr

Sennur Ulukus
Dept. of Electrical and Computer Engineering
University of Maryland
College Park, MD, USA
ulukus@umd.edu

***Abstract*—We study a pull-based monitoring system in which a common remote monitor queries the states of a collection of heterogeneous finite-state irreducible continuous time Markov chain (CTMC) based information sources, according to a Poisson process with different per-source sampling rates, in order to maintain remote estimates of the states. Three information freshness models are considered to quantify the accuracy of the remote estimates: fresh when equal (FWE), fresh when sampled (FWS) and fresh when close (FWC). For each of these freshness models, closed-form expressions are derived for mean information freshness for each source, as a function of the sampling rate. Using these expressions, optimum sampling rates for all sources are obtained using water-filling based optimization for maximizing the weighted sum freshness of the monitoring system, under an overall sampling rate constraint. Numerical examples are presented to validate the effectiveness of the proposed method by comparing it to several baseline sampling policies.**

I. Introduction

Timely delivery of status packets from a number of information sources for maintaining information freshness at a remote monitoring point has recently gained significant attention for the development of applications requiring real-time monitoring and control. In push-based systems, information sources decide when to sample and form the information packets, which are subsequently forwarded towards the destination [1], whereas in pull-based systems, monitors proactively query the information sources upon which sampling takes place [2].

In this paper, the interest will be on the remote estimation of a collection of finite, irreducible, time-homogeneous (transition probabilities do not change with time) CTMC based information sources for which the average rate at which state changes take place, are different among the sources. In some applications including cache update systems, the information at the destination may not hold a value unless the content at the destination is synchronized with the source. In this case, *binary freshness* process is suitable which takes the value of one when the information at the destination is up to date (or is in sync with the source), and is zero otherwise [3], [4].

In this paper, we consider the pull-based communication system in Fig. 1 in which N CTMC based information sources,

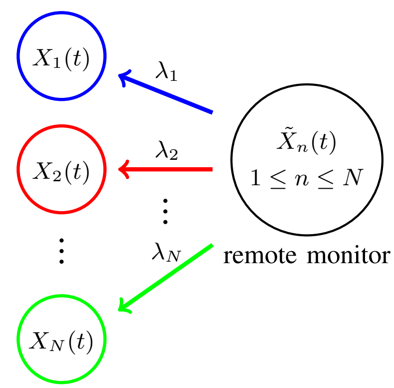

Fig. 1. Remote monitor queries the states of the CTMC $X_n(t)$ associated with source-n according to a Poisson process with rate λ_n to maintain an estimate $\tilde{X}_n(t)$ of $X_n(t)$. The dynamics of CTMC sources are assumed to be known by the monitor.

denoted by $X_n(t)$, $t \geq 0$, are sampled by a common remote monitor according to a Poisson process with intensity λ_n to maintain a martingale estimate $\tilde{X}(t)$, $t \geq 0$, of the original processes where the martingale estimate for the CTMC $X_n(t)$ at a future epoch, given the current and preceding observed states, is defined as the current observed state.

We propose three freshness models to quantify the accuracy of the martingale estimators. In the *fresh when equal* (FWE) binary freshness model, the information is fresh at the destination when the remote estimator is in sync with the state of the original source. In the *fresh when sampled* (FWS) freshness model (which is also binary), the information becomes stale when the state of the original source changes and it becomes fresh only when a new sample is taken. The non-binary *fresh when close* (FWC) model is a generalization of FWE, where the level of freshness depends on the proximity of the original process and its estimator at the monitor.

In this work, for the three freshness models of interest, we find the optimum sampling rate for each source that maximizes a weighted sum freshness of the system, under a total sampling rate constraint for the monitor. We show that this optimization problem possesses a water-filling structure; see [5] for a survey on water-filling based optimization for wireless networks. Iterative methods are available for water-filling based optimization which will be shown to give rise to

This work is done when N. Akar is on sabbatical leave at University of Maryland, MD, USA, as a visiting professor, which is supported in part by the Scientific and Technological Research Council of Turkey (Tübitak) 2219-International Postdoctoral Research Fellowship Program.

an algorithm for the optimum monitoring problem of Fig. 1 with quadratic worst-case computational complexity in the number of sources.

The contributions of our paper are as follows: 1) To the best of our knowledge, optimum sampling of heterogeneous (state-space and transition rates are different across the sources) CTMCs under overall sampling rate constraints has not been explored in the literature. 2) We derive closed-form expressions for mean freshness for FWE, FWC, and FWS models for finite, irreducible CTMCs. For the FWS model, the obtained expression is in terms of the sum of first order rational functions of the sampling rate which is a strictly concave increasing function. For the FWE and FWC models, similar expressions are obtained for the sub-case of time-reversible CTMCs. 3) The obtained expressions allow us to use computationally efficient water-filling algorithms and obtain optimum sampling policies.

II. Related Work

The remote estimation problem of information sources is studied in various works. The reference [6] investigates the problem of sampling a Wiener process with samples forwarded to a remote estimator over a channel that is modeled as a queue. The authors of [7] study a transmitter monitoring the evolution of a two-state discrete Markov source and sending status updates to a destination over an unreliable wireless channel for the purpose of real-time source reconstruction for remote actuation. The work presented in [8] studies the trade-off between the sampling frequency and staleness in detecting the events through a freshness metric called age penalty which is defined as the time elapsed since the first transition out of the most recently observed state. The authors of [9] investigate the effect of AoI on the accuracy of a remote monitoring system. A common feature of the above works is the existence of a single information source which gets to be sampled. On the other hand, the authors of [10] study the sampling of a collection of heterogeneous two-state CTMC-based information sources each modeling whether an individual is infected with a virus or not while using the binary freshness metric, and [11] studies sampling of multiple heterogeneous Poisson processes representing the citation indices of multiple researchers with the goal of keeping timely estimates of all the random processes. In the current paper, we study a heterogeneous collection of general finite-state CTMC-based information sources with binary freshness and/or binary-state CTMCs being sub-cases.

III. System Model

We consider the monitoring system in Fig. 1 with N continuous-time information sources each of which being a finite-state, irreducible CTMC. The CTMC associated with source-n is denoted by $X_n(t)$, $n \in \mathcal{N} = \{1, 2, \ldots, N\}$, $t \geq 0$, and $X_n(t) \in \{1, 2, \ldots, K_n\}$, where K_n is the size of the state space for $X_n(t)$. The process $X_n(t)$ has the infinitesimal generator Q_n of size K_n with its (i, j)th entry denoted by $q_{n,ij}$ which is the transition rate from state i to state j for $i \neq j$ and its diagonal entries are strictly negative satisfying $Q_n e = 0$, where e is a column vector of ones of appropriate size. Q_n has a left eigenvector π_n and right eigenvector e, associated with the simple eigenvalue of zero and all other eigenvalues having strictly negative real parts [12]. The ith element of the steady-state vector π_n is denoted by $\pi_{n,i}$ which is the steady-state probability of the process $X_n(t)$ being in state i, i.e., $\pi_{n,i} = \lim_{t\to\infty} \mathbb{P}[X_n(t) = i]$. The steady-state vector satisfies $\pi_n Q_n = 0$ and $\pi_n e = 1$. The transition rate out of state i is denoted by $\sigma_{n,i} = \sum_{j\neq i} q_{n,ij}$. The average transition intensity of source-n is denoted by r_n, i.e., $r_n = \sum_{i=1}^{K_n} \pi_{n,i}\sigma_{n,i}$ which is the long-term frequency of state transitions for the CTMC $X_n(t)$. Intuitively, sources with larger transition intensities need to be sampled at higher sampling rates for keeping the information fresh at the remote monitor; in this paper, we will determine exactly how often each source should be sampled relative to others, depending on the given system parameters. We denote by r the system transition intensity, $r = \sum_{i=1}^{N} r_n$.

The remote monitor in Fig. 1 queries the original process $X_n(t)$ according to a Poisson process with intensity $\lambda_n > 0$ in order to maintain an estimate of the instantaneous state of the original information process. In this paper, we propose to use the martingale estimator $\tilde{X}_n(t)$ which is given as $\tilde{X}_n(t) = X_n(t')$, where t' is the latest sampling time before t. The martingale estimator is very simple to maintain, and moreover, we will show that it is possible to obtain closed-form expressions for the performance of martingale estimators for time-reversible CTMCs, which subsequently leads to an algorithmic optimization procedure for obtaining the sampling rates in a heterogeneous setting.

The accuracy of the remote estimator is studied with three information freshness models described below. For the FWE information freshness model, the information is said to be fresh at the remote monitor only when the original process and its estimate are equal, i.e., $\tilde{X}_n(t) = X_n(t)$, and is otherwise stale. In the FWE model, there is no value at all in a sample, unless the original process and its estimator are synchronized. Hence, the binary freshness process $F_{n,e}(t)$ is defined as $F_{n,e}(t) = 1$ when $\tilde{X}_n(t) = X_n(t)$, and $F_{n,e}(t) = 0$ otherwise. Irreducible, finite-state CTMCs are ergodic. Therefore, mean freshness, $\mathbb{E}[F_{n,e}]$, which is written in terms of the limiting behavior of $F_{n,e}(t)$, is also equal to the time-averaged freshness,

$$\mathbb{E}[F_{n,e}] = \lim_{t\to\infty} \mathbb{P}[F_{n,e}(t) = 1], \tag{1}$$

$$= \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau F_{n,e}(t)\,\mathrm{d}t\,. \tag{2}$$

In the FWC model, we assume that there may be a value in a sample when the original process and its estimator are close enough to each other despite being out of sync, from a certain semantic perspective. For this purpose, for FWC, we introduce a proximity matrix $P_n = \{p_{n,ij}\}$, $0 \leq p_{n,ij} \leq 1$ for source-n, and subsequently define a non-binary freshness process $F_{n,c}(t)$ which takes the value $p_{n,ij}$ when $X_n(t) = i$ and $\tilde{X}_n(t) = j$. In particular, $p_{n,ii} = 1$, representing perfect freshness when the original process and its estimator are synchronized. Close to

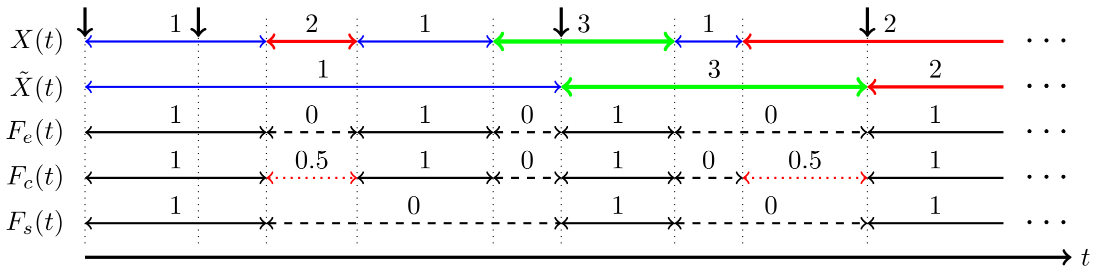

Fig. 2. Sample paths of the processes $X(t)$, $\tilde{X}(t)$, $F_e(t)$, $F_c(t)$ and $F_s(t)$ for a source with 3 states for an example scenario. Down arrows represent sampling instances.

unity values of $p_{n,ij}$ are representative of proximity between the states i and j. When P_n is taken as the identity matrix, FWC reduces to FWE.

For the FWS model, the binary freshness process $F_{n,s}(t)$ is set whenever the process is sampled, and it stays set until $X_n(t)$ makes a transition at which instant $F_{n,s}(t)$ becomes zero. For FWS, when the freshness process is stale, it cannot be set until the remote monitor re-samples the process, which is in contrast to FWE. The advantage of the FWS model is that it represents certain applications where freshness can be regained only with an explicit user action once it has been lost, and further, closed from expressions can be obtained for FWS for a very general class of CTMCs making it suitable for convex optimization in these general cases.

The mean freshness metrics $\mathbb{E}[F_{n,c}]$ and $\mathbb{E}[F_{n,s}]$ for the FWC and FWS models, respectively, are defined similar to (2). When the index of the source is immaterial, the subscript n is dropped for the source process $X(t)$ and its estimator $\tilde{X}(t)$ along with the corresponding freshness processes $F_e(t)$, $F_c(t)$, and $F_s(t)$ for FWE, FWC, and FWS, respectively.

Fig. 2 depicts the sample paths of the processes $X(t)$, $\tilde{X}(t)$, $F_e(t)$, $F_c(t)$ and $F_s(t)$ for a source with three states $\{1,2,3\}$ for an example scenario where the proximity matrix P for FWC is chosen such that $p_{ij}=1$ when $|i-j|=0$, 0.5 when $|i-j|=1$, and zero when $|i-j|=2$. Note that when $F_s(t)=1$, then $F_e(t)=1$, but not otherwise. Therefore, $\mathbb{E}[F_e]\geq\mathbb{E}[F_s]$ for any choice of the sampling rate λ. Moreover, $\mathbb{E}[F_c]\geq\mathbb{E}[F_e]$ since $F_c(t)$ can be larger than zero when $F_e(t)=0$, stemming from the structure of the proximity matrix P. Thus, $\mathbb{E}[F_c]\geq\mathbb{E}[F_e]\geq\mathbb{E}[F_s]$.

IV. Analytical Expressions for Mean Freshness

A. FWE Model

Theorem 1 provides an expression for $f(\lambda)=\mathbb{E}[F_e]$ for the FWE model for the CTMC $X(t)$.

Theorem 1. *Let the irreducible CTMC $X(t)\in\{1,2,\ldots,K\}$ with generator Q and steady-state vector π, be Poisson sampled with sampling rate λ. Then, for the FWE model, the mean freshness $f(\lambda)=\mathbb{E}[F_e]$ is given by,*

$$f(\lambda)=\lambda\pi\ \boldsymbol{diag}[(\lambda I-Q)^{-1}], \tag{3}$$

where $\boldsymbol{diag}[\cdot]$ represents a column vector composed of the diagonal entries of its matrix argument.

Proof. Let us consider the two-dimensional random process $Y(t)=(\tilde{X}(t),X(t))$, which is also Markov. To see this, note that, the transition intensity from state (i,j) to (i,j') is $q_{jj'}$ and from state (i,j) to (j,j) for $j\neq i$ is λ. Let the steady-state vector of the process $Y(t)$ be denoted by y, i.e., $y_{ij}=\lim_{t\to\infty}\mathbb{P}[Y(t)=(i,j)]$, $1\leq i,j\leq K$. Let Y be a $K\times K$ matrix such that $Y=\{y_{ij}\}$. The global balance equations (GBE) for CTMCs are a set of equations, one for each state s of the CTMC, which states that the total probability flux out of a state s should be equal to the total probability flux from other states into the state s, in steady-state. Applying the GBE for the state (i,i) of $Y(t)$ provides the following equation for each i, $1\leq i\leq K$,

$$y_{ii}\sigma_i=\sum_{j\neq i}y_{ij}q_{ji}+\sum_{j\neq i}y_{ji}\lambda, \tag{4}$$

$$=\sum_{j\neq i}y_{ij}q_{ji}+\lambda(\pi_i-y_{ii}), \tag{5}$$

where the last equality stems from the identity $\pi_i=\sum_j y_{ji}$. On the other hand, when the GBE is applied for the state (i,k), $k\neq i$, then we obtain the following,

$$y_{ik}(\sigma_k+\lambda)=\sum_{j\neq k}y_{ij}q_{jk},\quad 1\leq i\leq K,\ k\neq i. \tag{6}$$

Writing the equations (5) and (6) in a matrix form, we obtain for each i, $1\leq i\leq K$,

$$Y(i,:)(Q-\lambda I)=-\lambda\pi_i I(i,:), \tag{7}$$

where $Y(i,:)$ and $I(i,:)$ denote the ith row of Y and the ith row of the identity matrix, respectively. In FWE, the freshness process $F_e(t)=1$ when the joint process $Y(t)$ is visiting state (i,i) in the steady-state for some state i and $F_e(t)=0$ otherwise. Therefore, $\mathbb{E}[F_e]=\sum_{i=1}^{K}y_{ii}$, which results in the identity (3). $\square$

We now shift our focus to the special case of time-reversible CTMCs which subsume birth-death CTMCs as a subcase, for which it is not only possible to simplify, but also to show the strict concavity of, the expression in (3) with respect to the variable λ. A time-reversible CTMC has a generator Q which satisfies (see [12]), $\pi_i q_{ij}=\pi_j q_{ji},\ i\neq j$. Let $\Pi=\mathrm{diag}\{\pi_1,\pi_2,\ldots,\pi_K\}$, be the diagonal matrix composed of the entries of π. Also let $S=\Pi^{1/2}Q\Pi^{-1/2}$, which is a symmetric matrix. Symmetric matrices have real eigenvalues and they are diagonalizable by orthogonal transformations. Therefore, there exists an orthonormal matrix U such that $U^T SU = D = \text{diag}\{-d_1, -d_2, \ldots, -d_{K-1}, 0\}$, where $-d_i$, with $d_i > 0$, are the corresponding real eigenvalues of the matrix S. Moreover, the matrix defined by $T = \Pi^{-1/2}U$ diagonalizes the original generator Q, i.e., $T^{-1}QT = D$, and the eigenvalues of Q being the same as those of S. Let the (i, j)th entries of T and $\tilde{T} = T^{-1}$ be denoted by t_{ij} and $\tilde{t}_{ij}$, respectively. The way the transformation matrix T is defined, the ith row of $\tilde{T}$, i.e., the ith left eigenvector of Q, is obtained by post-multiplying by Π the transpose of the ith column of T, i.e., the transpose of the ith right eigenvector of Q. Moreover, the row vector π is the Kth row of $\tilde{T}$ and e is the Kth column of T. The following corollary gives a simplified expression for the mean freshness in terms of the sum of first-order rational functions of the variable λ for time-reversible CTMCs.

Corollary 1. *Consider the process $X(t)$ of Theorem 1 with generator Q which is time-reversible and with diagonalizing transformation matrix T. Then, the mean freshness $f(\lambda) = \mathbb{E}[F_e]$ is given for the FWE model by*

$$f(\lambda) = \sum_{j=1}^{K} \pi_j^2 + \lambda \sum_{j=1}^{K-1} \frac{b_j}{\lambda + d_j}, \tag{8}$$

$$= 1 - \sum_{j=1}^{K-1} \frac{a_j}{\lambda + d_j}, \tag{9}$$

where $b_j = \sum_{i=1}^{K} \pi_i^2 t_{ij}^2 = \sum_{i=1}^{K} \tilde{t}_{ji}^2, a_j = b_j d_j$ for $1 \leq j \leq K-1$. Moreover, $f(\lambda)$ is increasing and strictly concave, and has a continuous derivative $f'(\lambda)$ with $\lim_{\lambda \to 0^+} f(\lambda) = \sum_{j=1}^{K} \pi_j^2$.

Proof. We first write the term $A = (\lambda I - Q)^{-1} = T(\lambda I - D)^{-1}T$ appearing in (3) as follows,

$$A = T \ \text{diag} \left\{ \frac{1}{\lambda + d_1}, \ldots, \frac{1}{\lambda + d_{K-1}}, \frac{1}{\lambda} \right\} \ T^{-1}. \tag{10}$$

Using (3) and (10), we have

$$\frac{f(\lambda)}{\lambda} = \sum_{i=1}^{K} \pi_i A_{ii} = \sum_{i=1}^{K} \pi_i \sum_{j=1}^{K} t_{ij} \tilde{t}_{ji} \frac{1}{\lambda + d_j}, \tag{11}$$

$$= \sum_{j=1}^{K-1} \underbrace{\sum_{i=1}^{K} \pi_i^2 t_{ij}^2}_{b_j} \frac{1}{\lambda + d_j} + \frac{1}{\lambda} \sum_{i=1}^{K} \pi_i^2, \tag{12}$$

$$= \sum_{j=1}^{K-1} \frac{b_j}{\lambda + d_j} + \frac{1}{\lambda} \sum_{i=1}^{K} \pi_i^2, \tag{13}$$

since $\tilde{t}_{ji} = \pi_i t_{ij}$ and also $\sum_{i=1}^{K} \pi_i t_{iK} \tilde{t}_{Ki} = \sum_{i=1}^{K} \pi_i^2$ by observing that $t_{iK} = 1$ and $\tilde{t}_{Ki} = \pi_i$. The result in (13) gives the desired expression in (8). Then, (9) follows directly from (13) and also the fact that $\lim_{\lambda \to \infty} f(\lambda) = 1$.

Moreover, the coefficients a_j and b_j are strictly positive since $\pi_i > 0$ and the entries of a column of T cannot be all zero. A first-order rational function of λ in the form $-a/(\lambda + d)$ is increasing and strictly concave for $a, d > 0$ and sums of concave functions are also increasing and strictly concave, completing the proof. The expression pertaining to $\lim_{\lambda \to 0^+} f(\lambda)$ immediately follows from (8). □

B. FWC Model

Theorem 2 provides an expression for $f(\lambda) = \mathbb{E}[F_c]$ for the FWC model in terms of the sum of first-order rational functions of the variable λ for time-reversible CTMCs. However, the proof is omitted due to page limitations and it is along the same lines of the proof of Theorem 1.

Theorem 2. *Consider the process $X(t)$ of Theorem 1 with generator Q which is time-reversible and with diagonalizing transformation matrix T. Then, the mean freshness $f(\lambda) = \mathbb{E}[F_c]$ is given for the FWC model by*

$$f(\lambda) = \pi \sum_{i=1}^{K} \pi_i P(:, i) + \sum_{j=1}^{K-1} \frac{b_j \lambda}{\lambda + d_j}, \tag{14}$$

$$= 1 - \sum_{j=1}^{K-1} \frac{a_j}{\lambda + d_j}, \tag{15}$$

where $b_j = \tilde{T}(j, :) \sum_{i=1}^{K} \pi_i t_{ij} P(:, i), a_j = b_j d_j, 1 \leq j \leq K-1$ and $P(:, i)$ denotes the i^{th} column of P, and $\lim_{\lambda \to 0^+} f(\lambda) = \pi \sum_{i=1}^{K} \pi_i P(:, i)$.

However, the coefficients b_j (and hence a_j) are not guaranteed to be non-negative.

C. FWS Model

Theorem 3 provides an expression for $f(\lambda) = \mathbb{E}[F_s]$ for the FWS model for the CTMC $X(t)$ whose proof is omitted due to page limitations.

Theorem 3. *Let the irreducible CTMC $X(t) \in \{1, 2, \ldots, K\}$ with generator Q and steady-state vector π be Poisson sampled with sampling rate λ. Then, for the FWS model, the mean freshness $f(\lambda) = \mathbb{E}[F_s]$ is given by,*

$$f(\lambda) = 1 - \sum_{i=1}^{K} \frac{a_i}{\lambda + \sigma_i}, \tag{16}$$

where $a_i = \pi_i \sigma_i$, $f(\lambda)$ is increasing and strictly concave, and has a continuous derivative $f'(\lambda)$ with $\lim_{\lambda \to 0^+} f(\lambda) = 0$.

V. Optimum Monitoring of Heterogeneous CTMCs

The monitor is resource-constrained, and therefore, there is a total sampling rate constraint Λ on the overall sampling rate of the monitor, i.e., $\lambda = \sum_{n=1}^{N} \lambda_n \leq \Lambda$.

Let us first focus our attention to the FWE freshness model for time-reversible CTMCs in which case we use the mean freshness metric $f_n(\lambda_n) = \mathbb{E}[F_{n,e}]$ for source-n, and the weighted sum freshness (or the system freshness) $F_S = \sum_{n=1}^{N} w_n f_n(\lambda_n)$, for the overall monitoring system where the normalized weights $w_n, n = 1, \ldots, N$, $\sum_n w_n = 1$, reflect the relative importance of the freshness of the information processes. Thus, we have the following optimization problem for weighted sum freshness maximization,

$$\begin{aligned} \max_{\lambda_n \geq 0} \quad & \sum_{n=1}^{N} w_n f_n(\lambda_n) = 1 - \sum_{n=1}^{N} \sum_{j=1}^{K_n - 1} \frac{w_n a_{n,j}}{\lambda_n + d_{n,j}} \\ \text{s.t.} \quad & \sum_{n=1}^{N} \lambda_n \leq \Lambda \end{aligned} \tag{17}$$

In (17), the coefficients $a_{n,j}, d_{n,j} > 0$, for $1 \leq j \leq K_n - 1$ are to be obtained for the CTMC $X_n(t)$ using the procedure described in Corollary 1 and the expression (9). The function $f_n(\lambda_n)$ is increasing and strictly concave, and has a continuous first order derivative $f_n'(\lambda_n)$ that monotonically decreases from the value ∞ at $\lambda_n = -d_n^*$ to zero as λ_n is increased to ∞, where $d_n^* = \min_j d_{n,j}$. This optimization problem is known to have a water-filling solution [5] on the basis of which Algorithm 1 provides an efficient solution to the optimization problem (17) which requires at most $N - 1$ iterations until termination. Step 2 of Algorithm 1 can be solved by using the two-dimensional bisection search algorithm detailed in [5].

The algorithm is outlined as follows. Initially, $I_n = 1$ for $n = 1, \ldots, N$. Then, for a given $\mu > 0$ and for each n such that $I_n = 1$, we iteratively find the value of $\lambda_n \in (-d_n^*, \infty)$ that satisfies $w_n f_n'(\lambda_n) = \mu$ using an inner bisection search algorithm. Once λ_n's are obtained, we check whether $\sum_{n=1}^{N} \lambda_n I_n < \Lambda$ or not, and we vary the value of μ according to an outer bisection search algorithm, which iteratively finds the value of μ such that $\sum_{n=1}^{N} \lambda_n I_n = \Lambda$. If $\lambda_n \leq 0$ at this step, then I_n and λ_n are set to zero for all such n and the procedure above is repeated.

For the special case of two-state CTMCs, i.e., $K_n = 2$, a closed-form solution is available for the solution of the equations in Step 2 since the inverse function of $f_n'(\cdot)$ can be written in closed form. In this case, it is not difficult to show that the choices of $\mu = \left(\frac{\sum_{n=1}^{N} \sqrt{w_{n,1} a_{n,1} I_n}}{\Lambda + \sum_{n=1}^{N} d_{n,1} I_n} \right)^2$ and $\lambda_n = \sqrt{\frac{w_n a_{n,1}}{\mu}} - d_{n,1}$ for sources with $I_n = 1$, provide a single-shot solution for Step 2 of Algorithm 1 without a requirement for bisection search for this step.

We note that, for the FWS model, Algorithm 1 can be used with the only difference being the upper limit of the inner summation changed to K_n in (17). For the FWC model, since the coefficients a_j's in (15) can be negative, concavity is not proven. However, we propose to use the same water-filling algorithm also for the FWC model based on the observation that the expression (15) turned out to be concave in all the examples we studied.

VI. Numerical Examples

In the first numerical example, we focus on FWE and FWS in a scenario of $N = 50$ heterogeneous two-state Markov chains with $\pi_{n,1} = 0.3$, $\pi_{n,2} = 0.7$ and linearly spaced transition intensities, i.e., $r_n = r_{n-1} + \delta$, $n = 2, \ldots, N$. In the numerical example, we set $r_1 = 0.01$ and the average transition intensity $\frac{1}{N} r = \frac{1}{N} \sum_{n=1}^{N} r_n$ is set to 10 which yields the choice of $\delta = 0.4078$. We denote by κ the sampling ratio which is defined as the ratio of the overall sampling rate Λ to the system transition intensity, i.e., $\kappa = \frac{\Lambda}{r}$. Obviously, the sampling ratio κ should be sufficiently large so as to keep the remote estimates of all the information sources fresh. The source weights are assumed to be the same with $w_n = \frac{1}{N}$. Algorithm 1 is used for both FWE and FWS models to obtain the optimum sampling rates λ_n's under an overall sampling rate constraint Λ which is chosen to attain a given sampling ratio κ.

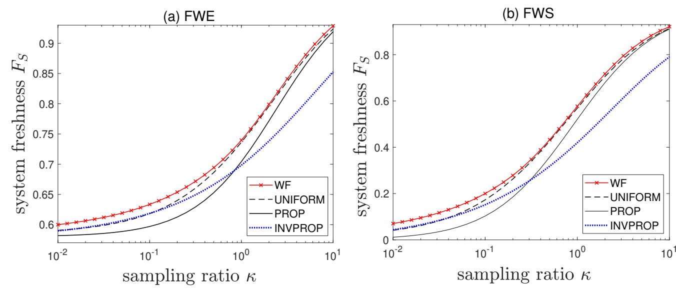

Fig. 3. The system freshness F_S as a function of the sampling ratio κ for the WF, UNIFORM, PROP, and INVPROP sampling policies: (a) FWE (b) FWS.

We compare our proposed water-willing solution, namely WF, with three baseline policies: i) UNIFORM policy samples each source-n uniformly likely, i.e., $\lambda_n = \frac{\Lambda}{N}$, ii) PROP policy chooses the sampling rate λ_n proportional with the source's transition intensity r_n, i.e., $\lambda_n \propto r_n$, iii) INVPROP policy chooses the sampling rate λ_n inversely proportional with the source's transition intensity r_n, i.e., $\lambda_n \propto \frac{1}{r_n}$. Fig. 3 depicts the system freshness F_S as a function of the sampling ratio κ when the WF, UNIF, PROP, and INVPROP sampling policies are employed for both FWE and FWS freshness models. For both models, we observe that the WF policy outperforms all the other three baseline policies. The UNIFORM policy yields very close to optimum freshness performance when the sampling ratio increases. However, for low sampling ratios, it is substantially outperformed by the WF policy. The PROP and INVPROP sampling policies perform poorly for small and large sampling ratios, respectively, against all other policies.

Fig. 4 depicts the optimum sampling rate λ_n divided by

Algorithm 1 Water-filling algorithm for the optimization problem (17) based on [5].

Step 1: Initialize $I_n = 1$ for $n = 1, \ldots, N$.
Step 2: Solve the following equations for the water level $\mu > 0$ and λ_n when $I_n = 1$,

$$w_n f_n'(\lambda_n) = \mu, \ \sum_{n=1}^{N} \lambda_n I_n = \Lambda. \tag{18}$$

Step 3: If $\lambda_n > 0$ for all n such that $I_n = 1$, then terminate while returning λ_n's.
Step 4: Otherwise, set $I_n = 0$ and $\lambda_n = 0$ for all n such that $I_n = 1$ and $\lambda_n \leq 0$, and go to Step 2.

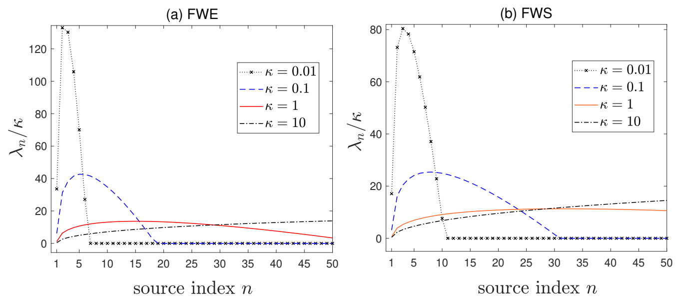

Fig. 4. The optimum sampling rate λ_n divided by κ as a function of the source index n for four values of the sampling ratio κ: (a) FWE (b) FWS.

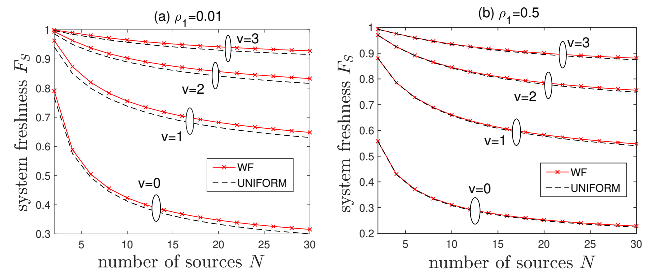

Fig. 5. System freshness F_S for the FWC model as a function of the number of users, N, for the WF and UNIFORM sampling policies for various values of the proximity parameter v.

the sampling ratio κ as a function of the source index n for different values of the sampling ratio κ for both freshness models. We observe that, for low sampling ratios, the water-filling solution chooses not to sample at all, a portion of the sources with high transition intensities, for both models. However, when the sampling ratio is sufficiently high, e.g., $\kappa = 10$, the optimum sampling rate for a given source appears to be increasing with the transition intensity of the source. Although the general behavior of the optimum sampling rate with respect to source index is quite similar for FWE and FWS models, in the latter model, the optimum sampling rate is more uniform across the sources for FWS than FWE.

In the final example, we focus on the FWC model and the choice of the proximity matrix P in terms of a proximity parameter v such that $p_{ij} = 1$ when $|i - j| \leq v$, and is zero otherwise. It is clear that FWC with $v = 0$ reduces to FWE. We consider an independent collection of N CTMCs each of which corresponds to the number of active servers in a multi-server M/M/c/c queuing system with $c = 10$ servers, with common service rate γ, and arrival rate ξ_n for source-n. The load for source-n is denoted by $\rho_n = \xi_n / c\gamma$. In this example, we assume linearly spaced loads, $\rho_n = \rho_{n-1} + \delta$, $n = 2, \ldots, N$ and the parameter δ is chosen so that the average load is fixed to $\rho_{avg} = \frac{1}{N}\sum_{n=1}^{N} \rho_n$. In this example, we fix $\gamma = 1$ and $\rho_{avg} = 0.9$. The source weights are identical as in the previous examples. The overall sampling rate bound is taken as $\Lambda = 20$. The weighted sum freshness $F_S = \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}[F_{n,c}]$ is plotted in Fig. 5 as a function of the number of users N for FWC with the proximity parameter $0 \leq v \leq 3$ and for two values of ρ_1 for the WF and UNIFORM policies. We have the following observations: When ρ_1 is close to ρ_{avg}, all the sources have similar statistical behaviors and therefore the performances of WF and UNIFORM policies should be similar which is evident from Fig. 5(b). However, WF substantially outperforms UNIFORM sampling policy in Fig. 5(a) where the smallest load source-1 has a load $\rho_1 = 0.01$ and consequently the sources are statistically dissimilar from each other. We have observed similar gains with WF over UNIFORM policy for all values of v we investigated.

VII. Conclusions

We studied a pull-based monitoring system which samples a heterogeneous collection of CTMCs according to a Poisson process, under an overall sampling rate constraint, employing a remote martingale estimate of the states of each of the CTMCs. Expressions for mean freshness are obtained for three freshness models of interest, and the optimum sampling rates for all CTMCs are obtained using water-filling based optimization while maximizing the weighted sum freshness. The worst case computational complexity of the proposed method is quadratic in the number of CTMCs making it possible to solve for scenarios even with very large number of CTMCs. The optimum monitoring policy is shown to outperform a number of heuristic baseline policies especially when there is diversity in the statistical characteristics of the underlying sources. Future work will consist of studying estimators other than the martingale estimator, and push-based systems.

References

[1] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1807–1827, March 2019.

[2] F. Li, Y. Sang, Z. Liu, B. Li, H. Wu, and B. Ji, "Waiting but not aging: Optimizing information freshness under the pull model," *IEEE/ACM Transactions on Networking*, vol. 29, no. 1, pp. 465–478, February 2021.

[3] M. Bastopcu and S. Ulukus, "Maximizing information freshness in caching systems with limited cache storage capacity," in *Asilomar Conference on Signals, Systems, and Computers*, November 2020.

[4] ——, "Information freshness in cache updating systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1861–1874, March 2021.

[5] C. Xing, Y. Jing, S. Wang, S. Ma, and H. V. Poor, "New viewpoint and algorithms for water-filling solutions in wireless communications," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1618–1634, February 2020.

[6] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the Wiener process for remote estimation over a channel with random delay," *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 1118–1135, February 2020.

[7] N. Pappas and M. Kountouris, "Goal-oriented communication for real-time tracking in autonomous systems," in *IEEE ICAS*, August 2021.

[8] J. P. Champati, M. Skoglund, M. Jansson, and J. Gross, "Detecting state transitions of a Markov source: Sampling frequency and age trade-off," *IEEE Transactions on Communications*, vol. 70, no. 5, pp. 3081–3095, May 2022.

[9] Y. Inoue and T. Takine, "AoI perspective on the accuracy of monitoring systems for continuous-time Markovian sources," in *IEEE Infocom*, April 2019.

[10] M. Bastopcu and S. Ulukus, "Using timeliness in tracking infections," *Entropy*, vol. 24, no. 6, p. 779, May 2022.

[11] ——, "Who should Google Scholar update more often?" in *IEEE Infocom*, July 2020.

[12] R. G. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013.