Information Aging through Queues: A Mutual Information Perspective

Yin Sun and Benjamin Cyr

Dept. of ECE, Auburn University, Auburn, AL

Abstract—In this paper, we propose a new measure for the freshness of information, which uses the mutual information between the real-time source value and the delivered samples at the receiver to quantify the freshness of the information contained in the delivered samples. Hence, the "aging" of the received information can be interpreted as a procedure that the above mutual information reduces as the age grows. In addition, we consider a sampling problem, where samples of a Markov source are taken and sent through a queue to the receiver. In order to optimize the freshness of information, we study the optimal sampling policy that maximizes the time-average expected mutual information. We prove that the optimal sampling policy is a threshold policy and find the optimal threshold exactly. Specifically, a new sample is taken once a conditional mutual information reduces to a threshold, and the threshold is equal to the optimum value of the time-average expected mutual information that is being maximized. Numerical results are provided to compare different sampling policies.

I. Introduction

Information usually has the greatest value when it is fresh [1]. For example, real-time knowledge about the location, orientation, and speed of motor vehicles is imperative in autonomous driving, and the access to timely updates about the stock price and interest-rate movements is essential for developing trading strategies on the stock market. In [2], [3], the concept of *Age of Information* was introduced to measure the freshness of information that a receiver has about the status of a remote source. Consider a sequence of source samples that are sent through a queue to a receiver, as illustrated in Fig. 1. Each sample is stamped with its generation time. Let U_n be the time stamp of the newest sample that has been delivered to the receiver by time instant n. The age of information, as a function of n, is defined as $\Delta_n = n - U_n$, which is the time elapsed since the newest sample was generated. Hence, a small age Δ_n indicates that there exists a fresh source sample at the receiver.

In practice, different sources may change at diverse speeds. For example, the location of a car can change much faster than the temperature of its engine. While the age of information Δ_n represents the time difference between the samples available at the transmitter and receiver, it is independent of the changing speed of the source. Hence, the age Δ_n is not an appropriate measure for comparing the freshness of information for different sources.
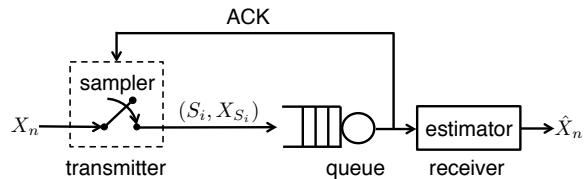
Fig. 1. System model.

In recent years, several examples and approaches for evaluating the freshness of information about time-correlated sources have been discussed in, e.g., [4]–[13]. In [4]–[6] and the references therein, the received samples are used to estimate the source value in real-time, where the estimation error is used to measure the freshness of information available at the receiver. In [7], an age penalty function $p(\Delta)$ was employed to describe the level of dissatisfaction for having aged samples at the receiver, where p is an *arbitrary non-negative* and *non-decreasing function* of the age Δ that can be specified based on the application; in addition, an optimal sampling strategy was developed to minimize the time-average expected age penalty function. In [8], the authors considered the relationship between the auto-correlation function $r(\Delta_n) = \mathbb{E}[X_n^* X_{n-\Delta_n}]$ (where X_n denotes the source status at time instant n) and the age penalty function in [7], and provided analytical expressions for the long-run time average of a few auto-correlation functions. In [9]–[13], several scheduling policies were developed to minimize an *arbitrary non-decreasing functional* $f(\{\Delta_n : n \geq 0\})$ of the age process $\{\Delta_n : n \geq 0\}$ in several network settings. The age penalty models in [9]–[13] are quite general, which include most age penalty models considered in previous studies as special cases. For example, because the functional $f(\{\Delta_n : n \geq 0\})$ is a mapping from the space of age processes to real numbers, it can be selected to describe the time-average age (i.e., $1/N \sum_{n=0}^{N} \Delta_n$), or the time-average of an age penalty function that depends on the age levels at multiple time instants (i.e., $1/N \sum_{n=0}^{N} p(\Delta_n, \Delta_{n-1}, \ldots, \Delta_{n-k})$).

In this paper, we propose a new measure for the freshness of information. In particular, the mutual information between the real-time source value and the received samples is used to quantify the freshness of the information contained in the delivered samples. This mutual information term is easy to compute for Markov sources and can precisely describe how the information ages over time. The detailed contributions of

this paper are summarized as follows:

- By using the data processing inequality, we show that for Markov sources, the above mutual information term is a *non-negative* and *non-increasing* function of the age Δ_n (Lemma 1). Therefore, the "aging" of the received information can be interpreted as a procedure that this mutual information reduces as the age Δ_n grows.

- In order to optimize the freshness of information, we study the optimal sampling strategy that maximizes the time-average expected mutual information. This problem is solved in two steps: *(i)* We generalize [7] to obtain an optimal sampling strategy that minimizes the time-average expected age penalty function $\limsup_{N\to\infty} \frac{1}{N} \mathbb{E}[\sum_{n=1}^{N} p(\Delta_n)]$, where $p(\Delta)$ is an *arbitrary non-decreasing* function of the age Δ (Theorem 1). *(ii)* Next, we apply the result of Step *(i)* to a special age penalty function, i.e., the negative of the mutual information, which is a *non-positive* and *non-decreasing* function of the age.

- The obtained optimal sampling strategy has a nice structure: A new sample is taken once a conditional mutual information reduces to a threshold β, and the threshold β is equal to the optimum value of the time-average expected mutual information that we are maximizing (Theorem 2). Numerical results are provided to compare different sampling policies.

A. Relationship with Previous Work

The closest study to this paper is [7]. The differences between [7] and this paper are explained in the following:

- As we have mentioned, the age penalty function $p(\cdot)$ in [7] is *non-negative* and *non-decreasing*, which is relaxed to be an *arbitrary non-decreasing* function in this paper.

- In [7], a two-layered nested bisection search algorithm was developed to compute the threshold β. In this paper, β is characterized as the solution of a fixed-point equation, which can be solved by a single layer of bisection search. Hence, the computation of β is simplified.

- In [7], the optimal sampling strategy was obtained for a continuous-time system. In this paper, we develop an optimal sampling strategy for a discrete-time system, without taking any approximation or sub-optimality.

- It was assume in [7] that after the previous sample was delivered, the next sample must be generated within a fixed amount of time. By adopting more powerful proof techniques, we are able to remove such an assumption and greatly simplify the proof procedure in this paper.

II. SYSTEM MODEL

We consider a discrete-time status-update system that is illustrated in Fig. 1, where samples of a source X_n are taken and sent to a receiver through a communication channel. The channel is modeled as a single-server FIFO queue with *i.i.d.* service times. The system starts to operate at time instant $n = 0$. The i-th sample is generated at time instant S_i and is delivered to the receiver at time instant D_i with a discrete

service time Y_i, where $S_1 \leq S_2 \leq \ldots$, $S_i + Y_i \leq D_i$, and $\mathbb{E}[Y_i] < \infty$ for all i. Each sample packet contains both the sampling time S_i and the sample value X_{S_i}. The samples that the receiver has received by time instant n are denoted by the set

$$\boldsymbol{W}_n = \{X_{S_i} : D_i \leq n\}. \tag{1}$$

At any time instant n, the receiver uses the received samples $\boldsymbol{W}_n$ to reconstruct an estimate $\hat{X}_n$ of the real-time source value X_n, where we assume that the estimator neglects the knowledge implied by the timing S_i for taking the samples.

Let $U_n = \max\{S_i : D_i \leq n\}$ be the time stamp of the freshest sample that the receiver has received by time instant n. Then, the *age of information*, or simply the *age*, at time instant n is defined as [2], [3]

$$\Delta_n = n - U_n = n - \max\{S_i : D_i \leq n\}. \tag{2}$$

The initial state of the system is assumed to satisfy $S_1 = 0$, $D_1 = Y_1$, and Δ_0 is a finite constant.

Let $\pi = (S_1, S_2, \ldots)$ represent a sampling policy and Π denote the set of *causal* sampling policies that satisfy the following two conditions: *(i)* Each sampling time S_i is chosen based on history and current information of the system, but not on any future information. *(ii)* The inter-sampling times $\{T_i = S_{i+1} - S_i, i = 1, 2, \ldots\}$ form a regenerative process [14, Section 6.1][1]: There exists an increasing sequence $0 \leq k_1 < k_2 < \ldots$ of almost surely finite random integers such that the post-k_j process $\{T_{k_j+i}, i = 1, 2, \ldots\}$ has the same distribution as the post-k_1 process $\{T_{k_1+i}, i = 1, 2, \ldots\}$ and is independent of the pre-k_j process $\{T_i, i = 1, 2, \ldots, k_j - 1\}$; in addition, $0 < \mathbb{E}[S_{k_{j+1}} - S_{k_j}] < \infty$, $j = 1, 2, \ldots$

We assume that the Markov chain X_n and the service times Y_i are determined by two mutually independent external processes, which do not change according to the adopted sampling policy.

III. MUTUAL INFORMATION AS A MEASURE OF THE FRESHNESS OF INFORMATION

In this paper, we propose to use the mutual information

$$I(X_n; \boldsymbol{W}_n) = H(X_n) - H(X_n|\boldsymbol{W}_n) \tag{3}$$

as a metric for evaluating the freshness of information that is available at the receiver. In information theory, $I(X_n; \boldsymbol{W}_n)$ is the amount of information that the received samples $\boldsymbol{W}_n$ carries about the real-time source value X_n. When the received samples $\boldsymbol{W}_n$ are fresh, $I(X_n; \boldsymbol{W}_n)$ is close to $H(X_n)$; when the received samples $\boldsymbol{W}_n$ are obsolete, $I(X_n; \boldsymbol{W}_n)$ is almost 0. In addition, because $I(X_n; \boldsymbol{W}_n)$ has naturally incorporated the information structure of the source X_n, it can effectively characterize different sources with diverse changing speeds and compare their information freshness.

---

[1]We assume that T_i is a regenerative process because we will optimize $\liminf_{N\to\infty} \mathbb{E}[\sum_{n=1}^{N} I(X_n; \boldsymbol{W}_n)]/N$, but operationally a nicer objective function is $\liminf_{i\to\infty} \mathbb{E}[\sum_{n=0}^{D_i} I(X_n; \boldsymbol{W}_n)]/\mathbb{E}[D_i]$. These two objective functions are equivalent if $\{T_1, T_2, \ldots\}$ is a regenerative process.

One interpretation of $I(X_n; \boldsymbol{W}_n)$ can be obtained by considering how helpful the received samples $\boldsymbol{W}_n$ are for predicting X_n. By using Shannon code lengths [15, Section 5.4], the expected minimum number of bits L required to specify X_n satisfies

$$H(X_n) \leq L < H(X_n) + 1. \tag{4}$$

Note that L is also the expected minimum number of binary experiments that are needed to infer X_n. If X_n is a random vector consisting of many symbols (e.g., X_n represents an image containing many pixels or the channel coefficients of many OFDM subcarriers), the one bit of overhead in (4) is not significant. With the knowledge of $\boldsymbol{W}_n$, the minimum number of bits required to specify X_n is at least $H(X_n|\boldsymbol{W}_n)$. Hence, $I(X_n; \boldsymbol{W}_n)$ is approximately the reduction in the description cost of predicting X_n without and with the knowledge of $\boldsymbol{W}_n$.

A. Markov Sources

Next, we consider the class of Markov sources and use the Markov property to simplify $I(X_n; \boldsymbol{W}_n)$. By using the data processing inequality [15], it is not hard to show that $I(X_n; \boldsymbol{W}_n)$ has the following property:

Lemma 1. *If X_n is a time-homogeneous Markov chain and $\boldsymbol{W}_n$ is defined in* (1)*, then the mutual information*

$$I(X_n; \boldsymbol{W}_n) = I(X_n; X_{n-\Delta_n}) \tag{5}$$

can be expressed as a non-negative and non-increasing function $r(\Delta_n)$ of the age Δ_n.

Proof. Because X_n is a Markov chain, $X_{\max\{S_i : D_i \leq n\}} = X_{n-\Delta_n}$ contains all the information in $\boldsymbol{W}_n = \{X_{S_i} : D_i \leq n\}$ about X_n. In other words, $X_{n-\Delta_n}$ is a sufficient statistic of $\boldsymbol{W}_n$ for estimating X_n. Then, (5) follows from [15, Eq. (2.124)].

Next, because X_n is time-homogeneous, $I(X_n; X_{n-\Delta}) = I(X_{\Delta+1}; X_1)$ for all n, which is a function of the Δ. Further, because X_n is a Markov chain, owing to the data processing inequality [15, Theorem 2.8.1], $I(X_{\Delta+1}; X_1)$ is non-increasing in Δ. Finally, $I(X_{\Delta+1}; X_1)$ is non-negative. This completes the proof. □

According to Lemma 1, information "aging" can be considered as a procedure that *the amount of information $I(X_n; \boldsymbol{W}_n)$ that is preserved in $\boldsymbol{W}_n$ for inferring the real-time source value X_n decreases as the age Δ_n grows*. This is similar to the data processing inequality [15] which states that no processing of the data Y can increase the information that Y contains about Z; the difference is that in the status-update systems that we consider, the data $\boldsymbol{W}_n$, the age Δ_n, and the concerned information X_n are all evolving over time.

Two examples of the Markov source X_n are provided in the sequel as illustrations of Lemma 1:

1) Gaussian Markov Source: Suppose that X_n is a first-order discrete-time Gaussian Markov process, defined by

$$X_n = a X_{n-1} + Z_n, \tag{6}$$

where $a \in (-1, 1)$ and the Z_n's are zero-mean *i.i.d.* Gaussian random variables with variance σ^2. Because X_n is a Gaussian Markov process, one can show that [16]

$$I(X_n; \boldsymbol{W}_n) = I(X_n; X_{n-\Delta_n}) = -\frac{1}{2} \log_2 \left(1 - a^{2\Delta_n}\right). \tag{7}$$

Since $a \in (-1, 1)$ and $\Delta_n \geq 0$ is an integer, $I(X_n; \boldsymbol{W}_n)$ is a positive and decreasing function of the age Δ_n. Note that if $\Delta_n = 0$, then $I(X_n; \boldsymbol{W}_n) = H(X_n) = \infty$, because the absolute entropy of a Gaussian random variable is infinite.

2) Binary Markov Source: Suppose that $X_n \in \{0, 1\}$ is a binary symmetric Markov chain defined by

$$X_n = X_{n-1} \oplus V_n, \tag{8}$$

where $\oplus$ denotes binary modulo-2 addition and the V_n's are *i.i.d.* Bernoulli random variables with mean $q \in [0, \frac{1}{2}]$. One can show that

$$I(X_n; \boldsymbol{W}_n) = I(X_n; X_{n-\Delta_n}) = 1 - H\left(\frac{1 - (1-2q)^{\Delta_n}}{2}\right), \tag{9}$$

where $\Pr[X_n = 1 | X_0 = 0] = \frac{1-(1-2q)^n}{2}$ and $H(x)$ is the binary entropy function defined by $H(x) = -x \log_2 x - (1 - x) \log_2(1-x)$ with a domain $x \in [0, 1]$ [15, Eq. (2.5)]. Because $H(x)$ is increasing on $[0, \frac{1}{2}]$, $I(X_n; \boldsymbol{W}_n)$ is a non-negative and decreasing function of the age Δ_n.

IV. ONLINE SAMPLING PROBLEM

To optimize the freshness of information, we formulate an online sampling problem for maximizing the time-average expected mutual information between X_n and $\boldsymbol{W}_n$ over an infinite time-horizon:

$$\bar{I}_{\text{opt}} = \sup_{\pi \in \Pi} \ \liminf_{N \to \infty} \frac{1}{N} \ \mathbb{E}\left[\sum_{n=1}^{N} I(X_n; \boldsymbol{W}_n)\right], \tag{10}$$

where $\bar{I}_{\text{opt}}$ is the optimal value of (10). We assume that $\bar{I}_{\text{opt}}$ is finite.

It is important to emphasize that $\bar{I}_{\text{opt}}$ in (10) is different from the Shannon capacity considered in, e.g., [15], [17]: In (10), our goal is to maximize *the freshness of information* and make more accurate inference about the real-time source value; this goal is achieved by minimizing the average amount of mutual information that is lost as the received data becomes obsolete. On the other hand, the focus of Shannon capacity theory is mainly on maximizing *the rate of information* that can be reliably transmitted to the receiver, but (in most cases) without significant concerns about whether the received information is new or old.

V. OPTIMAL ONLINE SAMPLING POLICY

In [7], an age penalty function $p(\Delta)$ was defined to characterize the level of dissatisfaction for having aged information at the receiver, where $p : \mathbb{R} \mapsto \mathbb{R}$ is an arbitrary *non-negative* and *non-decreasing* function that can be specified according to the application. For continuous-time status-update systems, the optimal sampling policy for minimizing the time-average
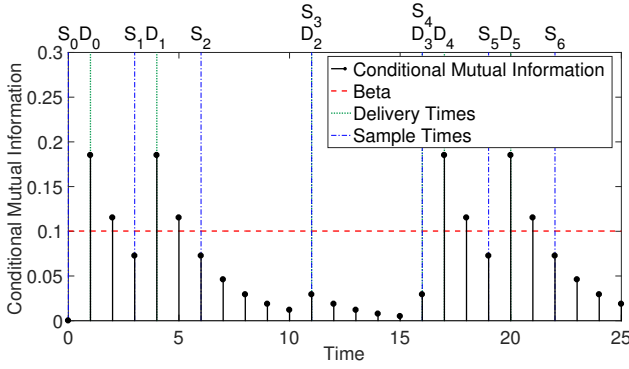
Fig. 2. A sample-path illustration of the optimal sampling policy (14) and (15), where the service time Y_i is equal to either 1 or 5 with equal probability. On this sample-path, the service times are $Y_0 = 1, Y_1 = 1, Y_2 = 5, Y_3 = 5, Y_4 = 1, Y_5 = 1, Y_6 = 5$.

expected age penalty $\limsup_{T\to\infty} \frac{1}{T} \mathbb{E}[\int_0^T p(\Delta(t))dt]$ was obtained in [7]. Unfortunately, we are not able to apply the results in [7] to solve (10). Specifically, if we choose an age penalty function $p_2(\Delta_n) = -I(X_n; W_n) = -r(\Delta_n)$, then Lemma 1 suggests that $p_2(\cdot)$ is a *non-positive* and *non-decreasing*, which is different from the *non-negative* and *non-decreasing* age penalty function required in [7]. In addition, we consider a discrete-time system in this paper, which is different from the continuous-time system in [7].

To address this problem, we generalize [7] by considering an arbitrary *non-decreasing* age penalty function (no matter *positive* or *negative*) and design an optimal sampling policy that minimizes the time-average expected age penalty. To that end, we consider the following discrete-time age penalty minimization problem:

$$\bar{p}_{\text{opt}} = \inf_{\pi\in\Pi} \limsup_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\sum_{n=1}^{N} p(\Delta_n)\right] \quad (11)$$

where $p : \mathbb{R} \mapsto \mathbb{R}$ is an arbitrary *non-decreasing* function and $\bar{p}_{\text{opt}}$ denotes the optimal value of (11). We assume that $\bar{p}_{\text{opt}}$ is finite. Problem (11) is a Markov decision problem. A closed-form solution of (11) is provided in the following theorem:

Theorem 1. *If $p : \mathbb{R} \mapsto \mathbb{R}$ in (11) is non-decreasing and the service times Y_i are i.i.d., then there exists a threshold $\beta \in \mathbb{R}$ such that the sampling policy*

$$S_{i+1} = \min\{n \in \mathbb{N}: n \geq D_i, \mathbb{E}[p(n + Y_{i+1} - S_i)|S_i] \geq \beta\} \quad (12)$$

is optimal to (11), where $D_i = S_i + Y_i$ and β is determined by solving (12) and (13):

$$\beta = \frac{\mathbb{E}\left[\sum_{n=D_i}^{D_{i+1}-1} p(n - S_i)\right]}{\mathbb{E}[D_{i+1} - D_i]}, \quad (13)$$

Further, β is exactly the optimal value of (11), i.e., $\beta = \bar{p}_{\text{opt}}$.

Proof Sketch. Theorem 1 is proven in three steps: *(i)* We show that it is better to generate samples when the server is idle.

This property helps to remove the harmful queueing delay and simplify the policy space containing the optimal policy. *(ii)* We use Lemma 2 of [18] to transform the problem into an equivalent problem that is easier to handle. *(iii)* Finally, we show that the optimal sampling problem can be decomposed into a sequence of mutually independent per-sample control problems. The optimal per-sample control policy is given by (12), with the threshold β satisfying (13). Due to space limitations, the details are relegated to our technical report [19]. $\square$

Next, we consider a special case that $p(\Delta_n) = -I(X_n; W_n) = -r(\Delta_n)$. It follows from Theorem 1 that

Theorem 2. *If the service times Y_i are i.i.d., then there exists a threshold $\beta \geq 0$ such that the sampling policy*

$$S_{i+1} = \min\{n \in \mathbb{N}: n \geq D_i,$$
$$\mathbb{E}_{Y_{i+1}}\left[I(X_{n+Y_{i+1}}; X_{S_i}|Y_{i+1} = y_{i+1})\right] \leq \beta\}$$
$$= \min\{n \in \mathbb{N}: n \geq D_i, I(X_{n+Y_{i+1}}; X_{S_i}|Y_{i+1}) \leq \beta\} \quad (14)$$

is optimal to (10), where $D_i = S_i + Y_i$, $\mathbb{E}_Y$ denotes the expectation with respect to the random variable Y, and β is determined by solving (14) and (15):

$$\beta = \frac{\mathbb{E}\left[\sum_{n=D_i}^{D_{i+1}-1} I(X_n; X_{S_i})\right]}{\mathbb{E}[D_{i+1} - D_i]}. \quad (15)$$

Further, β is exactly the optimal value of (10), i.e., $\beta = \bar{I}_{\text{opt}}$.

The optimal sampling policy in (14) and (15) has a nice structure: The next sampling time S_{i+1} is determined based on the mutual information between the freshest received sample X_{S_i} and the signal value $X_{D_{i+1}}$, where $D_{i+1} = S_{i+1} + Y_{i+1}$ is the delivery time of the $(i + 1)$-th sample. Because the transmission time Y_{i+1} will be known by both the transmitter and receiver at time $D_{i+1} = S_{i+1} + Y_{i+1}$, Y_{i+1} is the side information that is characterized by the conditional mutual information $I[X_{n+Y_{i+1}}; X_{S_i}|Y_{i+1}]$. The conditional mutual information $I[X_{n+Y_{i+1}}; X_{S_i}|Y_{i+1}]$ decreases as time n grows. According to (14), the $(i + 1)$-th sample is generated at the smallest integer time instant n satisfying two conditions: *(i)* The i-th sample has already been delivered, i.e., $n \geq D_i$, and *(ii)* The conditional mutual information $I[X_{n+Y_{i+1}}; X_{S_i}|Y_{i+1}]$ has reduced to be no greater than a pre-determined threshold β. One can observe from (14) that at any time n, there is at most one sampling in transmission and no sample is waiting in the queue. This is a big difference from traditional sampling policies such as uniform sampling, where the freshness of information is significantly affected by the queueing delay, as shown later in Fig. 3.

In addition, according to (15), the threshold β is equal to the optimum objective value $\bar{I}_{\text{opt}}$ in (10), i.e., the optimum of the time-average expected mutual information $\liminf_{N\to\infty} \frac{1}{N} \mathbb{E}[\sum_{n=1}^{N} I(X_n; W_n)]$ that we are maximizing. Note that the sampling times S_i and delivery times D_i on the
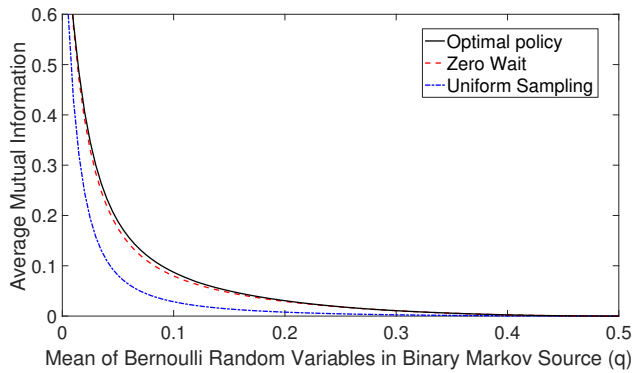
Fig. 3. Time-average expected mutual information vs the mean q of Bernoulli random variables W_n for the binary Markov source in (8).

right-hand side of (15) depends on β. Hence, β is a fixed point of (15).

The optimal sampling policy is illustrated in Fig. 2, where the service time Y_i is equal to either 1 or 5 with equal probability. The service time S_i, delivery time D_i, and conditional mutual information $I[X_{n+Y_{i+1}}; X_{S_i}|Y_{i+1}]$ of the samples are depicted in the figure. One can observed that if the service time of the previous sample is $Y_i = 1$, the sampler will wait until the conditional mutual information $I[X_{n+Y_{i+1}}; X_{S_i}|Y_{i+1}]$ drops below the threshold β and take the next sample then; if the service time of the previous sample is $Y_i = 5$, the next sample is taken at the delivery of the previous sample at time D_i, because the conditional mutual information $I[X_{n+Y_{i+1}}; X_{S_i}|Y_{i+1}]$ is already below β.

VI. NUMERICAL RESULTS

In this section, we evaluate the freshness of information achieved in the following three sampling policies:

- *Uniform sampling*: Periodic sampling with a period given by $S_{i+1} - S_i = \mathbb{E}[Y_i]$.
- *Zero-wait*: In this sampling policy, a new sample is taken once the previous sample is delivered to the receiver, so that $S_{i+1} = D_i = S_i + Y_i$.
- *Optimal policy*: The sampling policy given by Theorem 2.

Let I_{uniform}, $I_{\text{zero-wait}}$, and I_{opt} be the average mutual information of these three sampling policies.

We consider the binary Markov source X_n in (8). The service time Y_i is equal to either 1 or 11 with equal probability.[2] Figure 3 depicts the time-average expected mutual information versus the mean q of the Bernoulli random variables V_n in (8). One can observe that $I_{\text{opt}} \geq I_{\text{zero-wait}} \geq I_{\text{uniform}}$ holds for every value of q. Notice that because of queueing delay, I_{uniform} is much smaller than I_{opt} and $I_{\text{zero-wait}}$. In addition, as q grows from 0 to 0.5, the changing speed of the binary Markov source X_n increases and the freshness of information (i.e., the time-average expected mutual information) decreases. When $q = 0.5$, the X_n's form an *i.i.d.* sequence and the freshness of information is zero in all three sampling policies.

[2]The service time distribution is different from that used in Figure 2.

VII. CONCLUSION

In this paper, we have used mutual information to evaluate the freshness of the received samples that describe the status of a remote source. We have developed an optimal sampling policy that can maximize the time-average expectation of the above mutual information. This optimal sampling policy has been shown to have a nice structure. In addition, we have generalized [7] by finding the optimal sampling strategies for minimizing the time-average expectation of *arbitrary non-decreasing* age penalty functions. We will consider other related information freshness metrics in our future work.

REFERENCES

[1] C. Shapiro and H. Varian, *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press, 1999.
[2] X. Song and J. W. S. Liu, "Performance of multiversion concurrency control algorithms in maintaining temporal consistency," in *Fourteenth Annual International Computer Software and Applications Conference*, Oct 1990, pp. 132–139.
[3] S. Kaul, R. D. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *IEEE INFOCOM*, 2012.
[4] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *IEEE ISIT*, July 2012, pp. 2666–2670.
[5] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," in *IEEE ISIT*, 2017.
[6] X. Gao, E. Akyol, and T. Baar, "Optimal communication scheduling and remote estimation over an additive noise channel," *Automatica*, vol. 88, pp. 57 – 69, 2018.
[7] 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 Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.
[8] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case," in *IEEE ISIT*, June 2017, pp. 326–330.
[9] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *IEEE ISIT*, 2016.
[10] ——, "Age-optimal information updates in multihop networks," in *IEEE ISIT*, 2017.
[11] ——, "Minimizing the age of information through queues," submitted to *IEEE Trans. Inf. Theory*, 2017, http://arxiv.org/abs/1709.04956.
[12] ——, "The age of information in multihop networks," submitted to *IEEE Trans. Inf. Theory*, 2017, https://arxiv.org/abs/1712.10061.
[13] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, "Age-optimal updates of multiple information flows," in *IEEE INFOCOM Workshops — the 1st Workshop on the Age of Information (AoI Workshop)*, 2018.
[14] P. J. Haas, *Stochastic Petri Nets: Modelling, Stability, Simulation*. New York, NY: Springer New York, 2002.
[15] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.
[16] I. M. Gel'fand and A. M. Yaglom, "Calculation of the amount of information about a random function contained in another such function," *American Mathematical Society Translations*, vol. 12, pp. 199–246, 1959.
[17] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, Jan 1996.
[18] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," Jan. 2017, http://arxiv.org/abs/1701.06734.
[19] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," technical report, 2018, http://www.auburn.edu/%7eyzs0078/ageMI.pdf.