Real-Time Status Updates for Markov Source

Sudheer Poojary, *Member, IEEE*, Sanidhay Bhambay, *Student Member, IEEE*, and Parimal Parag, *Member, IEEE*

***Abstract*— For timely sensor update, the traditional approach is to send new information at every available opportunity. Recent research has shown that with limited receiver feedback, sensors can improve the update timeliness by transmitting differential information for slowly varying correlated sources. One can elect to transmit either the actual or the differential state information based on a *differential encoding threshold* for a general Markov source. This threshold captures the natural trade-off between differential transmission opportunities and the coding gains. Using *matrix-geometric method*, we find the limiting age distribution for a Markov source as a function of the encoding threshold, from which several other performance metrics of interest, such as mean age, peak age, and the probability of decoding failure can be derived.**

***Index Terms*— Age of information, Markov source, erasure channel, matrix-geometric method, differential encoding, block codes.**

I. Introduction

In THE past few years we have seen a rapid proliferation of 'connected' devices which aid in real time decision making. Many of these devices typically monitor physical phenomena such as temperature, pressure, humidity, traffic, pollution, etc. and communicate this information to the cloud to receive central-server-aided decisions. In these real-time actuation/decision systems, the timeliness of data is crucial. Data timeliness is also important in many other applications such as social media updates, distributed system updates, security patches, and route updates in ad-hoc networks.

Information age was introduced in [2], as a metric to measure information staleness. At any time t, the *age of information* can be defined in terms of the generation time $U(t)$ of the last received information symbol as

$$A(t) \triangleq t - U(t). \quad (1)$$

Lower information age at the receiver would correspond to a more timely information, and this is the metric we adopt in this article. Most of the existing literature on communication timeliness [3]–[7], study a queueing theoretic abstraction where the update generation and delivery times are stochastic. In abovementioned works, the update is always received correctly at the receiver, and the channel uncertainty is captured in the random reception time of an update.

In contrast to the typical queueing theoretic models, we model the channel unreliability by an information theoretic binary erasure channel. In this setup, a received update may not be decodable and can be dropped. This setup is different from a typical queueing setup, where an update would be received a random time after transmission and the updates are transmitted and received in order without any packet drops. Contrastingly in our model, one can choose not to re-transmit an update upon reception failure, and transmit an entirely new update packet. That is, our model captures the scenario where packets can be dropped if certain service constraints are violated. Apart from capturing the decoding uncertainty, a simple information theoretic abstraction of channel allows us to study the impact of coding on update timeliness. Channel coding for timely update over independent and identically distributed (*iid*) erasure channels is studied for always-on source in [8], [9], for slowly varying source in [10], for special Markov sources in [1]. We adopt a similar setup, where the source sends an update packet that includes coded status information bits.

For a temporally correlated source, one can transmit either the current state or its difference from the last correctly decoded state, depending on the number of bits needed to represent either information. A straightforward update scheme is to merely encode the current source state, and transmit at each available opportunity, referred to as *true update*. This scheme is agnostic to the differential encoding opportunities due to the source correlation, and hence serves as an upper bound on the information age. Indeed, it is shown in [10], that sending differential update reduces information age considerably for a highly correlated source, in the presence of limited receiver feedback. We will refer to this scheme as *differential update with feedback* or *incremental update* for brevity. Reduction in information staleness in this scheme results from the exploitation of temporal correlation across messages. This is an optimistic lower bound on the information age of the system, due to the underlying assumption that the source can always exploit the message correlation to efficiently encode the differential updates. In practice, temporal correlation can vary between two transmission opportunities for a general source, and the number of bits needed to represent differential message depends on the actual realization of the states.

The practice of using differential coding is quite common and is used in a wide variety of applications such as rsync [11],

Manuscript received February 16, 2018; revised February 27, 2019; accepted April 10, 2019. Date of publication May 2, 2019; date of current version August 16, 2019. This work was supported by the Science and Engineering Research Board (SERB) under Grant DSTO-1677. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the SERB. This work was presented in part at the 2017 IEEE ITW [1].

S. Poojary is with Qualcomm Inc., Bengaluru 560066, India (e-mail: spoojary@qti.qualcomm.com).

S. Bhambay is with Samsung Inc., Bengaluru 560055, India (e-mail: sanidhay@iisc.ac.in).

P. Parag is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru 560012, India (e-mail: parimal@iisc.ac.in).

Communicated by N. Devroye, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2019.2913856

HTTP [12] and version control [13]. Differential encoding is typically used to reduce traffic or storage requirements. In our case, differential encoding gives us opportunities to use better error protection and consequently improve the age metric. We note that it may not always be efficient to use differential coding and any sudden and large change may require us to send the actual state. We will consider this in our work and use a general incremental update scheme wherein we send the actual state when the change in state does not permit differential encoding. For a finite state source, we assume that the number of bits needed to represent any actual state is m. In this article, the source sends a differential update only when the differential information can be represented by k bits, for some $k \leqslant m$. We call this threshold k as the *differential encoding level*, a design parameter that can be chosen to minimize information staleness for a given channel. We consider a *generalized incremental update* scheme, where the source transmits a true update only in the following two cases. First, if the source is unable to encode a differential update in k bits. Second, if the receiver fails to decode an update. When the last correctly decoded state is i, the source can encode the differential information in k bits with certain *differential encoding probability* depending on the source state i. That is, there is uncertainty in differential encoding opportunities at correlated source. For a slowly varying source, one can almost surely always exploit the temporal correlation and hence the differential encoding probability is unity. It follows that, the generalized incremental update scheme generalizes the incremental update scheme.

In this article, we study sources that have a Markov state evolution. We capture the notion of temporal correlation between successive status updates by setting the transition probabilities of the associated Markov process appropriately. For example, for a source with higher temporal correlation, we set state transitions to be more likely to the neighboring states than to states farther off. At the other extreme, i.e., of no correlation, we have an *iid* source where the transition to different states is independent of how close or far they are to the current state. Thus we can classify a slowly evolving source, i.e., a source with higher temporal correlation as a highly correlated source as opposed to an independent source where the updates are independent and hence have no correlation.

We show that the highly correlated source that can always send differential updates [10], and uniformly correlated source where the incremental update probability is independent of the source state [1] are special cases. We further note the natural trade-off between coding opportunities and coding gain that determines the selection of differential encoding level k. In particular, if we increase the differential encoding level k, then the number of additional parity bits $n - k$ available to incremental update reduces as compared to the true update. On the other hand, if we reduce the level k, only a handful of state differences can be sent as a differential update. Both these scenarios limit the performance gain of incremental update over true updates.

A. Main Contribution

Our main contribution is to characterize the limiting age distribution for the generalized incremental update scheme for a general Markov source. It turns out that the standard renewal theory technique can't be applied in a straightforward manner for this case. We identify a minimal Markov chain that captures both the age and the source state at each update reception instant. The transition matrix for this Markov chain has a block structure, and hence the limiting distribution can be computed using matrix-geometric methods. Equilibrium distribution for this Markov chain admit a geometric solution in terms of a fundamental matrix, which is a solution of a matrix polynomial equation [14]–[16]. Fortunately, for the system model under consideration, the marginal limiting distribution of age admits a closed form solution in terms of the system parameters. From this closed form limiting distribution, we can compute additional performance metrics such as limiting values of mean age, peak age, decoding failure probability, decay-rate of age etc. We can recover the existing results for special sources, such as independent, highly correlated, and uniformly correlated sources. We illustrate the trade-off between additional coding gains and the coding opportunities for a general Markov source.

B. Literature Review

Mean information age is computed for status update queueing systems under various queueing disciplines, service distributions, and arrival distributions in [2], [5], [6], [17]. Status update systems have the flexibility in determining which update is transmitted to the receiver. It was shown in [18], [19], that replacing the waiting update by the latest update can reduce the information age for a single source queue with memoryless arrival and service, and buffer size 2. This result was generalized to multiple sources in [20], and with status deadlines in [3]. Scheduling problem for multiple source updates over a single wireless channel is considered in [21]. Status update transmissions over specific networks have also been studied, such as parallel networks [22], single-hop networks [23], and multi-hop networks [24].

For the status update systems, where there is uncertainty not only in the service time but also in the successful reception, there is a pertinent question of whether to re-transmit old status update or to send a new update. This problem is considered for random arrivals in [4], and always-on source in [8], [9], [25]. Multiple sources with parallel unreliable links is considered in [26], and a multiple access channel in [27]. Channel-aware source update is considered in [7], and channel-state update is considered in [28].

Optimal source sampling strategy to minimize the mean square estimation error for Weiner process is considered in [29], where the reception time for real-valued samples are random. Instead of the process estimation error, we are adopting the simple metric of average age in our paper. It has been shown in [7] that these two problems are equivalent when the sampling is independent of the observed Weiner process. For simplicity of analysis, we assume a fixed length encoding and opportunistic sampling, which implies periodic sampling and transmission in our setting of discrete-valued Markov source.

We note that the joint Markov process under consideration in this paper is a discrete-time version of the stochastic hybrid systems (SHS) [6], [30], [31]. We apply techniques from matrix-geometric methods, which are more specific to the block-partition structure of the transition matrix for the discrete-time Markov process under consideration.

II. System Model

In this section, we describe each component of the communication model as shown in Fig. 2, and transmission protocol in detail.

A. Source

We assume a correlated physical process taking discrete values from a finite field $\mathbb{F}_{2^m}$, such that the state $M(t)$ can be represented by m bits at every discrete time sample $t \in \mathbb{N}$. We assume that the sampled physical process $(M(t) \in \mathbb{F}_{2^m} : t \in \mathbb{N})$ is an irreducible and aperiodic discrete time Markov chain with transition probability matrix $\mathbf{Q}$, and hence is positive recurrent with a unique invariant distribution ν such that

$$\nu\mathbf{Q} = \nu, \qquad \langle \nu, \mathbf{1} \rangle = 1.$$

In the following section, we will see that the transmission opportunities for the source arise only periodically at instants $\{(j-1)n+1 : j \in \mathbb{N}\}$. Accordingly, we assume that the source samples the physical process only at these discrete instants. We denote the discrete-time discrete-valued sampled physical process at the source by

$$M_j \triangleq M((j-1)n+1), \;\; j \in \mathbb{N}.$$

We call the random sequence $(M_j : j \in \mathbb{N})$, the source process which is a Markov chain with the transition probability matrix $\mathbf{P} = \mathbf{Q}^n$, and the invariant distribution ν since $\nu\mathbf{P} = \nu$.

The difference between two consecutively transmitted source states is denoted by $\delta_j \triangleq M_j - M_{j-1} \in \mathbb{F}_{2^m}$. For a fixed differential encoding level k, the set of possible values for the difference δ_j to be represented by k bits is denoted by

$$\Delta_k \triangleq \{-2^{k-1}, \dots, 2^{k-1} - 1\}.$$

The event $\{\delta_j \in \Delta_k\}$ is referred to as *differential encoding success*. The probability of differential encoding success depends on the previous source state M_{j-1} and the differential encoding level k. Conditioned on the previous source state M_{j-1} being $i \in \mathbb{F}_{2^m}$, we denote the encoding success probability as $P_i(\Delta_k)$, which can be computed for the Markov source $(M_j : j \in \mathbb{N})$ as

$$P_i(\Delta_k) \triangleq \Pr(M_j - M_{j-1} \in \Delta_k | M_{j-1} = i) = \sum_{j-i \in \Delta_k} \mathbf{P}_{ij}.$$

B. Encoder

We assume that the first source message is the m-bit source state M_1. For each $j \geqslant 2$, the source message can either be the m-bit current state M_j or the difference δ_j of the current state M_j from the previously transmitted state M_{j-1}, if the difference can be represented by k-bits and M_{j-1} is successfully decoded at the receiver. We assume a fixed length permutation invariant coding for both type of source messages, such that each source message is encoded into an n-length codeword $X_j \triangleq (X_{j1}, \dots, X_{jn})$ that is transmitted at discrete instant $t = (j-1)n+1$. The encoded message corresponding to the true state M_j is called *true update* and the encoded message corresponding to the state difference δ_j is called the *incremental update*. We denote the indicator to the event of the jth encoded message being a differential update by γ_j, and denote $1 - \gamma_j$ by $\bar{\gamma}_j$, to write the number of information bits in the jth codeword as

$$r_j = k\gamma_j + m\bar{\gamma}_j, \;\; j \in \mathbb{N}.$$

It is possible to use a rateless code instead of the fixed-length block code. Though for many real-time practical systems, we envision selection of a finite and fixed block-length n. Further, the preemption of the current update to begin immediate transmission of a new update requires a bit-wise feedback. Thus, we stick to the finite block-length setting since it requires a simpler implementation with reduced feedback overhead.

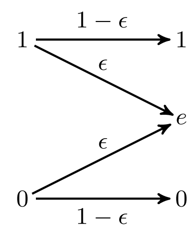

Fig. 1. A bit-wise *iid* binary symmetric erasure channel with erasure probability ϵ, where the erased bit is denoted by e.

C. Channel

We measure time-units in terms of channel uses, and assume that each bit-transmission requires single channel use. Therefore, an n-length update codeword requires n channel uses, and the jth codeword X_j transmitted at time $(j-1)n+1$ is received at instant $jn+1$. That is, the codeword length determines the periodicity of the source transmission opportunities.

We consider a bit-wise *iid* binary symmetric erasure channel, as depicted in Fig. 1. Over this channel, each transmitted bit can be successfully received, or erased independently and identically with probability ϵ. When the transmitted bit is erased, the received symbol is denoted by erasure symbol e. That is, corresponding to the channel input $X_{ji} \in \{0,1\}$ for the ith bit-transmission for the jth codeword, the channel output is denoted by $Y_{ji} \in \{0,1,e\}$ written as

$$Y_{ji} = e1_{\{Y_{ji} \neq X_{ji}\}} + X_{ji}1_{\{Y_{ji} = X_{ji}\}}, \;\; i \in [n].$$

Since the bit-wise erasure channel is *iid*, the erasure indicators $(1_{Y_{ji} \neq X_{ji}} : i \in [n], j \in \mathbb{N})$ are *iid* Bernoulli with mean $\Pr\{Y_{ji} \neq X_{ji}\} = \epsilon$. The number of erasures in the jth received codeword is sum of n independent erasure indicators,

$$E_j \triangleq \sum_{i \in [n]} 1_{\{Y_{ji} \neq X_{ji}\}}.$$

From the channel independence, it follows that the sequence $(E_j : j \in \mathbb{N})$ of number of erasures in received codewords is *iid* with the common Binomial distribution,

$$\Pr\{E_j = \ell\} = \binom{n}{\ell}\epsilon^{\ell}(1-\epsilon)^{n-\ell}, \text{ for } \ell \in \{0, 1, \dots, n\}.$$

D. Decoder

From the received channel output Y_j at time $t = jn+1$, the decoder computes an estimate $\hat{M}_j$ of the message transmitted n channel uses ago. For binary erasure channels, the receiver can either perfectly decode the transmitted message or declare a decoding failure. Thus, while the receiver may fail to decode an update, it will never admit an erroneous update. For the jth received update, the indicator to the event of decoding success is denoted by

$$\xi_j = 1_{\{M_j = \hat{M}_j\}}, \text{ for } j \in \mathbb{N}.$$

We call a block code to be *permutation invariant*, if the event of codeword decoding failure depends solely on the number of erasures E in the codeword, and not their locations. Writing the indicator of decoding failure as $\bar{\xi} = 1 - \xi$, we denote the probability of decoding failure for a permutation invariant code given E erasures in an n-length codeword with r information bits and $n - r$ parity bits by

$$P(n, n-r, E) = \mathbb{E}[\bar{\xi}|n-r, E].$$

The number of information bits in an n-length codeword is m for the true update, and k for the incremental update respectively. Hence, the unconditional probability of decoding failure for true update and incremental update is respectively, $p_a = \mathbb{E}[P(n, n-m, E)]$ and $p_d = \mathbb{E}[P(n, n-k, E)]$, where the expectation is taken over the binomial random variable E with parameters (n, ϵ). We assume erasure probability $\epsilon \in (0, 1)$, and hence $p_d, p_a \in (0, 1)$. Further, since the number of parity bits are higher for the differential update, we have $p_d < p_a$.

From *iid* nature of the channel, it follows that the codeword decoding failure events are also independent, conditioned on the differential update indicators $(\gamma_j : j \in \mathbb{N})$. The sequence of decoding failure indicators $(\bar{\xi}_j : j \in \mathbb{N})$ has conditional mean

$$\mathbb{E}[\bar{\xi}_j|\gamma_j] = \mathbb{E}[P(n, n-r_j, E_j)] = p_d\gamma_j + p_a\bar{\gamma}_j, \quad j \in \mathbb{N}. \quad (2)$$

E. Control Channel

We also assume the existence of a separate control channel [32], that allows the decoder to distinguish between a true and an incremental update. The system model is illustrated in Fig. 2.

III. Generalized Incremental Update

After the codeword reception, there are two possibilities at the receiver. First, the receiver is able to decode the transmitted message correctly, leading to a successful status update. Alternatively, the receiver declares a decoding failure and sends an immediate and accurate negative feedback to the source.[1] The source always responds to an event of decoding failure by sending its true state as the following update.

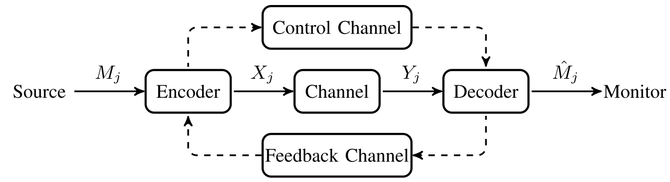

Fig. 2. An abstract discrete time communication model for a source, where the state M_j at time $t = (j-1)n + 1$ is encoded to an n-length codeword X_j, which is received as channel output Y_j after n channel uses at time $t = jn + 1$. Decoder forms an estimate $\hat{M}_j$ at the reception instant, where each transmitted bit is erased independently and identically.

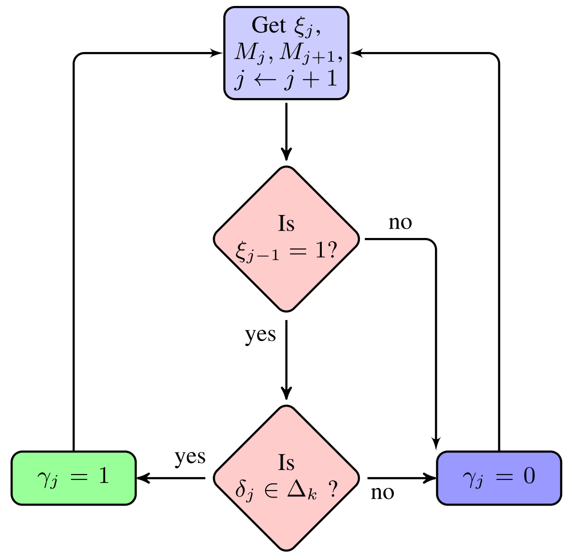

Fig. 3. We show the flow chart for the algorithm used by the source in generalized incremental update scheme, to encode message M_j as incremental or true update.

For the generalized incremental update, the source starts with a true update, encoding the first source state sample M_1. For any $j \geqslant 2$, the jth transmission is a differential update if and only if both of the following two conditions are met. First, the $(j-1)$th transmission was successfully received. Second, the jth state difference δ_j can be represented by k bits, i.e.

$$\gamma_j = \xi_{j-1}1_{\{\delta_j \in \Delta_k\}}, \quad j \geqslant 2. \quad (3)$$

This shows that the differential update indicator γ_j is function of M_j, M_{j-1}, ξ_{j-1}, and hence has a correlated evolution. Since the source Markov process $(M_j : j \in \mathbb{N})$ is independent of the decoding success sequence $(\xi_j : j \in \mathbb{N})$, we can find the mean of the differential update indicator γ_j conditioned on previous source state M_{j-1} and decoding success indicator ξ_{j-1} of previous update, as

$$\mathbb{E}[\gamma_j|\xi_{j-1}, M_{j-1}] = \xi_{j-1}P_{M_{j-1}}(\Delta_k).$$

We have summarized the generalized differential update algorithm at the source by a flow chart in Fig. 3.

[1]We note that, since we assume an erasure channel, the above two possibilities are the only ones that occur and the receiver never admits an erroneous update.

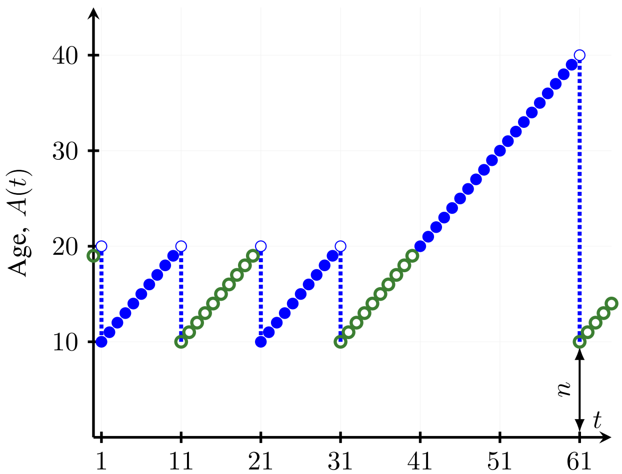

Fig. 4. This plot shows a sample path of the age process for the generalized incremental updates with codeword length $n = 10$. Hollow and filled parts of the age sample path correspond to bit-transmissions of incremental and true update codewords respectively. In this realization of age evolution, the first update at instant $t = 1$ is encoded as a true update and successfully decoded at instant $n+1$. Source fails to encode $M_3 - M_2$ within k bits, and sends the true update at time $2n + 1$ which is successfully decoded at instant $3n + 1$. Incremental update sent at instant $3n + 1$ fails to get decoded at instant $4n + 1$. Therefore, the source starts sending the true updates from the next transmission opportunity at instant $4n + 1$.

We observe that certain types of the probability transition matrix $\mathbf{P}$ reduces generalized incremental update to two special cases. For example, when $P_i(\Delta_k) = 0$ for each $i \in \mathbb{F}_{2^m}$, then the generalized scheme reduces to always sending true updates. Whereas, if $P_i(\Delta_k) = 1$ for each $i \in \mathbb{F}_{2^m}$, then the generalized scheme reduces to always sending incremental updates. Thus, the generalized update scheme generalizes both the previously studied update schemes in [1], [10].

IV. Age Process

Timeliness is our primary performance criterion, and we adopt the *age of information* [2] at the receiver as a measure of information staleness. In the following, we characterize the information age at receiver as a stochastic process. We focus on the sampled version of this process at update reception instants, which determines this process at all times.

Let $U(t)$ denote the generation time of the last correctly decoded update at the receiver at time t, then the age of information $A(t)$ at time t is given by $t - U(t)$. Since the source state is sampled at each transmission opportunity and instantaneously transmitted, the generation time of the last correctly decoded state is the corresponding transmission time. Hence, the generation time $U(t)$ remains constant until the reception of next successfully decoded update. Therefore, at each reception instant $t = jn + 1$ for $j \in \mathbb{N}$, we can write

$$U(t) = \xi_j(t - n) + \bar{\xi}_j U(t - n).$$

It follows that the age $A(t) = t - U(t)$ resets to value n at the instants of decoding success of update codewords, and is linearly increasing at all other instants. We have illustrated a sample path for the age process in Fig. 4. It follows that the age at the reception instants $(jn + 1 : j \in \mathbb{N})$ is a sequence of random variables of the form qn where $q \in \mathbb{N}$. Therefore, we consider the age process sampled at the jth reception instant $t = jn + 1$ and scaled by $1/n$, denoted by $A_j \triangleq A(jn + 1)/n$. The scaled and sampled age process is denoted by $(A_j : j \in \mathbb{N})$, and its evolution can be written in terms of decoding success indicators $(\xi_j : j \in \mathbb{N})$ as

$$A_j = 1 + \bar{\xi}_j A_{j-1}. \tag{4}$$

Assuming $A_0 \in \mathbb{N}$, it follows that $A_j \in \mathbb{N}$ for all $j \in \mathbb{N}$. In addition, the sampled-scaled age is unity if and only if the last reception was successful, that is $\{A_j = 1\}$ if and only if $\{\xi_j = 1\}$. We will show that the sampled-scaled age process is ergodic, and hence the limiting probability of sampled-scaled age can be written as

$$\lim_{j\to\infty} \Pr\{A_j \geqslant q\} = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} 1_{\{A_j \geqslant q\}}.$$

One can completely determine the age process at any discrete instant t from its sampled-scaled version. Since each update is sent using an n length codeword, the number of update receptions until a discrete time t can be defined as $N_t \triangleq \lfloor \frac{t-1}{n} \rfloor$, such that $nN_t \leqslant t - 1 < n(N_t + 1)$. We can write the age $A(t)$ at time t in terms of number of update receptions N_t as

$$A(t) = nA_{N_t} + (t - 1 - nN_t).$$

For example in Fig. 4, we have codeword length $n = 10$ and there are five updates received until time $t = 55$. It follows that at this time $t = 55$, the number of update receptions is $N_t = 5$, the sampled-scaled age at last update reception is $A_{N_t} = 3$, and the age is $A(55) = 34$. We will derive various performance metrics of interest from the limiting distribution of the sampled-scaled age process.

V. Joint Markov Process

The source is sampled at time $(j - 1)n + 1$ and its state M_j is encoded as jth source update, which is received at time $jn + 1$. The sampled-scaled age at the receiver, at the time of jth update reception is denoted by A_j. We study the joint evolution of the source state and the sampled-scaled age, and denote this process by the random sequence $((M_j, A_j) : j \in \mathbb{N})$. We say that the process is at level q if sampled-scaled age $A_j = q$.

We first show that the joint process $((M_j, A_j) \in \mathbb{F}_{2^m} \times \mathbb{N} : j \in \mathbb{N})$ is an irreducible and aperiodic Markov chain and determine the associated transition probability matrix. We further show that this joint Markov chain is positive recurrent and has a unique stationary distribution $\pi = (\pi_{i,q} : i \in \mathbb{F}_{2^m}, q \in \mathbb{N})$, where

$$\pi_{i,q} \triangleq \lim_{j\to\infty} \Pr\{(M_j, A_j) = (i, q)\}.$$

We can also write the invariant distribution as $\pi = (\pi_1, \pi_2, \dots)$, where $\pi_q = (\pi_{i,q} \in [0, 1] : i \in \mathbb{F}_{2^m})$ is the stationary distribution vector for the joint process to have age q. Age acts as level for the matrix-geometric methods described below, and we say that π_q is the invariant distribution vector for the joint process to be in level q.

The transition operator for the Markov process (M_j, A_j) has a block structure, and we characterize its invariant distribution using matrix-geometric methods [14]–[16]. Thereby, we derive the marginal distribution of the sampled-scaled age process by summing over all possible source states. From the invariant distribution of the sampled-scaled age, we can derive the performance metrics of interest.

Theorem 1. *The joint process of the source state and the sampled-scaled age* $((M_j, A_j) : j \in \mathbb{N})$ *is an irreducible, aperiodic, and positive recurrent homogeneous Markov chain with the transition probability operator* $\mathbf{T}$ *having the form*

$$\mathbf{T} = \begin{bmatrix} \tilde{\mathbf{B}} & \tilde{\mathbf{F}} & \mathbf{0} & \mathbf{0} & \dots & \dots \\ \mathbf{B} & \mathbf{0} & \mathbf{F} & \mathbf{0} & \dots & \dots \\ \mathbf{B} & \mathbf{0} & \mathbf{0} & \mathbf{F} & \dots & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix}, \tag{5}$$

where the $\mathbb{F}_{2^m} \times \mathbb{F}_{2^m}$ *matrices* $\tilde{\mathbf{B}}, \tilde{\mathbf{F}}, \mathbf{B}, \mathbf{F}$ *are given by*

$$\mathbf{B} = (1 - p_a)\mathbf{P}, \qquad \mathbf{F} = p_a\mathbf{P}$$
$$\tilde{\mathbf{B}} = \mathbf{B} + \tilde{\mathbf{D}}, \qquad \tilde{\mathbf{F}} = \mathbf{F} - \tilde{\mathbf{D}},$$

and the $\mathbb{F}_{2^m} \times \mathbb{F}_{2^m}$ *matrix* $\tilde{\mathbf{D}}$ *is defined by its* (i, l)*th entry*

$$\tilde{\mathbf{D}}_{il} = (p_a - p_d)\mathbf{P}_{il} 1_{\{l - i \in \Delta_k\}}.$$

Proof: State space of the process is obtained in Lemma 12. The process is shown to be Markov in Lemma 13, and homogeneous in Lemma 15. The corresponding transition probabilities are obtained in Corollary 16. The process is shown to be irreducible in Lemma 17, aperiodic in Lemma 18, and positive recurrent in Lemma 19. ∎

A. Matrix-Geometric Method

Any discrete-time Markov chain over countable state space is called a *quasi GI/M/1 process* [14]–[16] if the state space can be permuted to obtain the following block partitioned form of the associated transition matrix

$$\mathbf{T} = \begin{bmatrix} \tilde{\mathbf{Q}}_1 & \tilde{\mathbf{Q}}_0 & \mathbf{0} & \mathbf{0} & \dots & \dots \\ \tilde{\mathbf{Q}}_2 & \mathbf{Q}_1 & \mathbf{Q}_0 & \mathbf{0} & \dots & \dots \\ \tilde{\mathbf{Q}}_3 & \mathbf{Q}_2 & \mathbf{Q}_1 & \mathbf{Q}_0 & \dots & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix}.$$

Here, the matrices $(\mathbf{Q}_i : i \in \mathbb{Z}_+)$ and $(\tilde{\mathbf{Q}}_i : i \in \mathbb{Z}_+)$ are of identical finite size $N \times N$. A quasi GI/M/1 process can be considered to be a two dimensional Markov chain with countable state space $N \times \mathbb{Z}_+$. If the quasi GI/M/1 is in state (i, q) then it is in phase $i \in \{1, \dots, N\}$ and level $q \in \mathbb{Z}_+$.

From the repeated block structure of probability transition operator $\mathbf{T}$, it follows that the joint process (M_j, A_j) is a quasi GI/M/1 process, with the source state $M_j \in \{1, \dots, 2^m\}$ being the phase and the sampled-scaled age $A_j \in \mathbb{N}$ being the level. Further, in our setting $\tilde{\mathbf{Q}}_1 = \tilde{\mathbf{B}}$, $\tilde{\mathbf{Q}}_q = \mathbf{B}$ for $q \geqslant 2$, $\tilde{\mathbf{Q}}_0 = \tilde{\mathbf{F}}$, $\mathbf{Q}_q = \mathbf{0}$ for $q \in \mathbb{N}$, and $\mathbf{Q}_0 = \mathbf{F}$. In Fig. 5, we have shown the possible state transitions for this process. Each transition leads to either a unit increase in level, or resetting of the age level to unity. Each level has all possible source states indicating a vector of states at each level, and hence there is a matrix of possible transitions between two levels. Forward transition matrix corresponding to unit level increase is denoted by $\mathbf{F}$ for levels $q > 1$, and $\tilde{\mathbf{F}}$ for level $q = 1$. Similarly, backward transition matrix associated with the age resetting to unity is denoted by $\mathbf{B}$ for levels $q > 1$, and $\tilde{\mathbf{B}}$ for level $q = 1$.

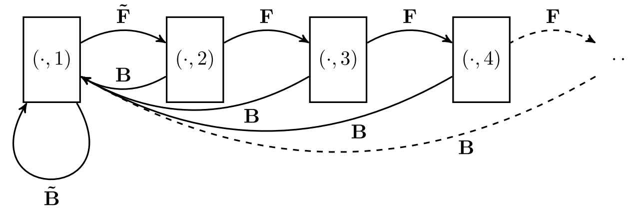

Fig. 5. We have shown the forward and backward block transitions for the joint process $((M_j, A_j) : j \in \mathbb{N})$. Each block of this Markov chain corresponds to a specific level of the joint process.

Since the backward transition matrix $\mathbf{B} = (1 - p_a)\mathbf{P}$ and the forward transition matrix $\mathbf{F} = p_a\mathbf{P}$, it follows that ν is the left eigenvector and $\mathbf{1}$ is a right eigenvector for both the matrices with the eigenvalue $(1 - p_a)$ and p_a respectively.

Next, we will show the relation between the stationary distribution ν of the sampled Markov source and the stationary distribution π of the joint process.

Lemma 2. *The following relation holds between the stationary distribution* ν *of the source state Markov chain* $(M_j : j \in \mathbb{N})$*, and the joint process* $((M_j, A_j) \in \mathbb{F}_{2^m} \times \mathbb{N} : j \in \mathbb{N})$*,*

$$\nu = \sum_{q \in \mathbb{N}} \pi_q.$$

Proof: Since the equilibrium distribution for the Markov chain $(M_j : j \in \mathbb{N})$ is ν, the result follows from the application of the monotone convergence theorem. Specifically, we have

$$\nu_i = \lim_{j \to \infty} \Pr\{M_j = i\} = \lim_{j \to \infty} \sum_{q \in \mathbb{N}} \Pr\{(M_j, A_j) = (i, q)\}$$
$$= \sum_{q \in \mathbb{N}} \lim_{j \to \infty} \Pr\{(M_j, A_j) = (i, q)\} = \sum_{q \in \mathbb{N}} \pi_{i,q}.$$

∎

Since the joint process is positive recurrent, the unique invariant distribution π satisfies the equilibrium condition $\pi\mathbf{T} = \pi$, for the associated transition probability operator $\mathbf{T}$. Next, we will find the equilibrium distribution of the joint process.

Theorem 3. *The equilibrium distribution* π *of the joint process* $((M_j, A_j) : j \in \mathbb{N})$ *with the associated transition probability operator* $\mathbf{T}$ *defined in* (5) *has the following geometric form*

$$\pi_q = \pi_1 \tilde{\mathbf{F}}\mathbf{F}^{q-2} \text{ for } q \geqslant 2, \quad \text{and } \pi_1 = (1 - p_a)\nu(\mathbf{I} - \tilde{\mathbf{D}})^{-1}.$$

Proof: From the equilibrium condition $\pi\mathbf{T} = \pi$ for the stationary distribution π, and the fact that $\tilde{\mathbf{B}} = \mathbf{B} + \tilde{\mathbf{D}}$, we get

$$\pi_1(\mathbf{I} - \tilde{\mathbf{D}}) = \sum_{q \in \mathbb{N}} \pi_q \mathbf{B}, \quad \pi_2 = \pi_1 \tilde{\mathbf{F}}, \ \pi_{q+1} = \pi_q \mathbf{F}, \ q \geqslant 2.$$

The geometric form of the level distribution follows from the second and the third equation above. Further, we notice that $\tilde{\mathbf{D}}$ is a sub-stochastic matrix with the spectral radius smaller than unity, and hence $\mathbf{I} - \tilde{\mathbf{D}}$ is invertible. The result follows from the first equation, using the result $\sum_{q\in\mathbb{N}} \pi_q = \nu$ from Lemma 2 and the fact that ν is a left eigenvector of $\mathbf{B}$ with eigenvalue $(1-p_a)$. ■

Corollary 4. *The sampled-scaled age process $(A_j : j \in \mathbb{N})$ is ergodic, with the stationary distribution for $q \geqslant 2$ given by*

$$\lim_{j\to\infty} \Pr\{A_j = q\} = \langle \pi_q, \mathbf{1}\rangle = (1-p_a)p_a^{q-2}(1 - \langle \pi_1, \mathbf{1}\rangle).$$

Proof: Ergodicity of the sampled-scaled age process follows from the ergodicity of the joint Markov chain, and we can find the associated marginal stationary distribution by summing over all source states. The result follows from the fact that $\mathbf{1}$ is a right eigenvector for $\mathbf{F}$ with eigenvalue p_a, $\langle \nu, \mathbf{1}\rangle = 1$, and that we can write $\tilde{\mathbf{F}} = (\mathbf{I} - \tilde{\mathbf{D}}) - (\mathbf{I} - \mathbf{F})$. ■

Lemma 5. *The sampled-scaled age process $(A_j : j \in \mathbb{N})$ is uniformly integrable, and hence it converges in mean. That is, $\lim_{j\to\infty} \mathbb{E}A_j = \mathbb{E}\lim_{j\to\infty} A_j$.*

Proof: Since the event $\{A_j = l\}$ implies decoding failure of last $l-1$ updates, we have

$$1_{\{A_j = l\}} = \prod_{i=0}^{l-2} \bar{\xi}_{j-i}\xi_{j-l+1} \leqslant \prod_{i=0}^{l-2} \bar{\xi}_{j-i}, \quad j \geqslant l.$$

Since the decoding failure probabilities for both true and differential updates are upper bounded by p_a, we can bound the following expectation for $j \geqslant q$ by

$$\begin{aligned}\mathbb{E}[|A_j| 1_{\{A_j \geqslant q\}}] &= \sum_{l\geqslant q} \mathbb{E}[|A_j| 1_{\{A_j = l\}}] \\ &\leqslant \sum_{l \geqslant q} l p_a^{l-1} = \frac{p_a^{q-1}(q(1-p_a) + p_a)}{(1-p_a)^2}.\end{aligned}$$

As $\lim_{q\to\infty} q p_a^{q-1} = 0$ and $\lim_{q\to\infty} p_a^q = 0$ for $p_a \in (0,1)$, we have $\lim_{q\to\infty} \sup_{j\in\mathbb{N}} \mathbb{E}[|A_j| 1_{\{A_j \geqslant q\}}] = 0$, and the uniform integrability follows. From the convergence in distribution for the sampled-scaled age and uniform integrability, the convergence in mean follows. ■

VI. Performance Evaluation

We compute the marginal limiting age distribution under the generalized incremental update scheme for a given Markov source on finite states $\mathbb{F}_{2^m}$ with transition matrix $\mathbf{Q}$, a fixed update codeword length n, and a fixed differential encoding threshold k. From the resulting marginal limiting age distribution, we evaluate the performance of the proposed generalized incremental update scheme. We denote the limit of sampled-scaled age by $A \triangleq \lim_{j\to\infty} A_j$, where the convergence is in distribution. Then,

$$\Pr\{A = q\} = \langle \pi_q, \mathbf{1}\rangle.$$

One can compute the moments of the limiting age process using the discrete Fourier transform. We define the z-transform of the stationary age distribution $(\langle \pi_1, \mathbf{1}\rangle, \langle \pi_2, \mathbf{1}\rangle, \dots)$ as

$$\Phi(z) \triangleq \sum_{q\in\mathbb{N}} z^q \langle \pi_q, \mathbf{1}\rangle. \qquad (6)$$

Using the explicit geometric from of the level distributions $(\pi_q : q \in \mathbb{N})$ and linearity of the dot product, we can write the z-transform of the equilibrium age distribution

$$\Phi(z) = \frac{(1-p_a)z^2}{1 - zp_a} + \frac{z(1-z)}{1-zp_a}\langle \pi_1, \mathbf{1}\rangle,$$

where the region of convergence is $|z| < 1/p_a$. We derive the following four performance metrics from the limiting age distribution.

A. Mean Information Age

First metric is the limiting empirical average age defined as $\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} A(s)$. We can compute this metric from the limiting distribution of sampled-scaled age, given by the following Lemma. The limiting sampled-scaled age is defined as $\lim_{j\to\infty} A_j$, where we take the convergence in distribution. Thus, we can formally define the distribution of limiting sampled-scaled age (also referred by limiting distribution of sampled-scaled age) as $\lim_{j\to\infty} \Pr\{A_j = q\}$. From the ergodicity of the sampled age process, it follows that the limiting empirical average of sampled-scaled age is equal to its limiting mean. Since the number of receptions until time t is $N_t = \lfloor (t-1)/n \rfloor$, we can write

$$\lim_{N_t\to\infty} \frac{1}{N_t}\sum_{j=1}^{N_t} A_j = \lim_{j\to\infty} \mathbb{E}[A_j].$$

From the uniform integrability of sampled-scaled age process shown in Lemma 5, we have the convergence of the sampled-scaled age process in the mean. In particular, the limiting mean of sampled-scaled age is equal to the mean of limiting sampled-scaled age.

Lemma 6. *The limiting average age is almost surely an affine function of limiting average of sampled-scaled age,*

$$\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} A(s) = n\mathbb{E}[A] + \frac{(n-1)}{2}.$$

Proof: We can express the cumulative sum of age between two reception instants $jn+1$ and $(j+1)n+1$ in terms of the sampled-scaled age at the jth reception, as $\sum_{s=jn+1}^{(j+1)n} A(s) = n\big(nA_j + \frac{(n-1)}{2}\big)$. Summing over all $s \in [t]$, we can write in terms of the number of receptions N_t until time t, as

$$n\sum_{j=1}^{N_t}\Big(nA_j + \frac{(n-1)}{2}\Big) \leqslant \sum_{s=1}^{t} A(s) \leqslant n\sum_{j=1}^{N_t+1}\Big(nA_j + \frac{(n-1)}{2}\Big).$$

Dividing by t and taking limits, it follows that the difference in limiting averages of the age and the sampled age is a constant,

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t} A(s) = \lim_{N_t\to\infty}\frac{1}{N_t}\sum_{j=1}^{N_t} nA_j + \frac{(n-1)}{2}.$$

From the ergodicity of the sampled-scaled age process shown in Corollary 4, it follows that the limiting empirical average of sampled-scaled age equals its limiting mean, that is $\lim_{N_t\to\infty}\frac{1}{N_t}\sum_{j=1}^{N_t}A_j = \lim_{j\to\infty}\mathbb{E}[A_j]$ almost surely. From Lemma 5, it follows that $\mathbb{E}[\lim_{j\to\infty}A_j] = \lim_{j\to\infty}\mathbb{E}[A_j]$. ■

Theorem 7. *The limiting average of sampled-scaled age for a Markov source with the generalized incremental update is*

$$\mathbb{E}[A] = \langle -\nu(\mathbf{I}-\tilde{\mathbf{D}})^{-1}, \mathbf{1}\rangle + 1 + \frac{1}{1-p_a}. \quad (7)$$

Proof: We can find the mean of the limiting sampled-scaled age as $\mathbb{E}A = \Phi'(1)$. The evaluation of first derivative of the z-transform, gives us

$$\Phi'(z) = (1-p_a)\frac{z(2-zp_a)}{(1-zp_a)^2} + \frac{(1-2z)+z^2p_a}{(1-zp_a)^2}\langle \pi_1, \mathbf{1}\rangle.$$

Result follows from the above by setting $z=1$, and substituting $\pi_1 = (1-p_a)\nu(\mathbf{I}-\tilde{\mathbf{D}})^{-1}$. ■

B. Limiting Probability of Decoding Failure

Second metric of interest is limiting probability of decoding failure $\lim_{j\to\infty}\Pr\{A_j \neq 1\}$. Since the sampled-scaled age process is ergodic, the limiting probability of decoding failure is almost surely equal to the limiting empirical average of number of decoding failures. That is,

$$\lim_{j\to\infty}\Pr\{A_j\neq 1\} = \lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}1_{\{A_j\neq 1\}} = \lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}\bar{\xi}_j.$$

It follows that the limiting probability of decoding failure is equal to limiting average of number of negative feedback messages from the receiver.

Theorem 8. *The limiting probability of decoding failure with generalized differential updates is*

$$\lim_{j\to\infty}\Pr\{A_j\geqslant 2\} = 1-\langle\pi_1,\mathbf{1}\rangle = 1-(1-p_a)\langle\nu(\mathbf{I}-\tilde{\mathbf{D}})^{-1},\mathbf{1}\rangle.$$

Proof: Follows from the marginal stationary distribution of sampled-scaled age in Corollary 4. ■

C. Tail Decay-Rate for Limiting Distribution

Third metric of interest is tail decay-rate or large deviation exponent of the limiting distribution for sampled-scaled age, defined as

$$\theta \triangleq -\lim_{q\to\infty}\frac{1}{q}\log\lim_{j\to\infty}\Pr\{A_j\geqslant q\}. \quad (8)$$

We can see from Corollary 4, that the limiting marginal distribution of the sampled-scaled age is geometrically distributed for $q\geqslant 2$, and hence the limiting distribution of sampled-scaled age decays with this large-deviation exponent θ even for finite threshold q. Thus, this exponent is an important performance metric of interest.

Theorem 9. *The tail decay-rate of limiting age distribution for Markov sources with generalized differential updates is*

$$\theta = -\lim_{q\to\infty}\frac{1}{q}\log\sum_{u=q}^{\infty}\langle\pi_u,\mathbf{1}\rangle = \log\frac{1}{p_a}. \quad (9)$$

Proof: From Corollary 4, we get that for $q\geqslant 2$

$$-\log\sum_{u=q}^{\infty}\langle\pi_u,\mathbf{1}\rangle = (q-2)\log\frac{1}{p_a} - \log(1-\langle\pi_1,\mathbf{1}\rangle).$$

The result follows from the definition of the tail decay-rate (8). ■

D. Distribution of Peak Information Age

Fourth metric is the limiting distribution of peak age [18], where the peak age is the age just before an update decoding success. That is, if $jn+1$ is an instant of update decoding success, then the peak age is $A(jn)$. We observe that we can write

$$A(jn) = A((j-1)n+1)+n-1 = nA_{j-1}+n-1.$$

We observe that the peak age is just a constant $n-1$ shift from the previous sampled age $A((j-1)n+1) = nA_{j-1}$, and hence we focus on finding the limiting distribution of the peak of the sampled-scaled age.

Peaks of the sampled-scaled age are equal to the number of update transmissions between two decoding successes. We denote the limiting sampled-scaled peak age of the generalized incremental update for Markov sources by A_M. The sampled-scaled age is unity if and only if the received update was successfully decoded. Hence, we can write the limiting distribution of the peak of sampled-scaled age as

$$\Pr\{A_M = q\} = \lim_{j\to\infty}\Pr\{A_j = q | A_{j+1}=1\}.$$

Theorem 10. *The limiting distribution of the peak of sampled-scaled age for Markov sources with the generalized incremental updates, is given by*

$$\Pr\{A_M=q\} = \begin{cases} 1-(1-p_a)\frac{(1-\langle\pi_1,\mathbf{1}\rangle)}{\langle\pi_1,\mathbf{1}\rangle} & q=1,\\ (1-p_a)^2p_a^{q-2}\frac{(1-\langle\pi_1,\mathbf{1}\rangle)}{\langle\pi_1,\mathbf{1}\rangle} & q\geqslant 2.\end{cases} \quad (10)$$

Proof: We can write the stationary distribution of joint process for $q\geqslant 2$ as

$$\lim_{j\to\infty}\Pr\{M_j=i, A_j=q, M_{j+1}=l, A_{j+1}=1\} = \pi_{i,q}\mathbf{B}_{il}.$$

By summing over source states M_j, M_{j+1}, we get

$$\lim_{j\to\infty}\Pr\{A_j=q, A_{j+1}=1\} = \langle\pi_q\mathbf{B},\mathbf{1}\rangle = (1-p_a)\langle\pi_q,\mathbf{1}\rangle.$$

The result for $q\geqslant 2$ follows from substituting the limiting marginal distribution of sampled-scaled age from Corollary 4 and the definition of conditional probability. The result for $q=1$ follows from the fact that $\Pr\{A_M=1\} = 1-\sum_{q\geqslant 2}\Pr\{A_M=q\}$. ■

Recall that the peak of sampled-scaled age is the number of updates between two decoding successes, and it can be seen that the limiting probability of decoding failure for generalized incremental updates of Markov sources, is given by

$$\lim_{j\to\infty} \Pr\{A_j \geqslant 2\} = 1 - \frac{1}{\mathbb{E}[A_M]}.$$

This relation can be established from equation (10) to compute the mean peak age as $\mathbb{E}[A_M] = \sum_{q\in\mathbb{N}} \Pr\{A_M \geqslant q\} = \frac{1}{\langle \pi_1, \mathbf{1}\rangle}$, and from Corollary 4 to get the limiting probability of decoding failure as $\lim_{j\to\infty} \Pr\{A_j \geqslant 2\} = 1 - \langle \pi_1, \mathbf{1}\rangle$.

VII. An Example Markov Source

We evaluate the performance metrics for the following Markov source, as a special case of our general result. Recall that ν denotes the invariant distribution of the transition probability matrix $\mathbf{P}$ of the Markov source. It follows that ν and all one column vector $\mathbf{1}$ are respectively the left and the right eigenvectors of the matrix $\mathbf{P}$ with unit eigenvalue.

A Markov source is called uniform if the probability of consecutive state difference δ_j being represented by k bits is independent of the initial state i. That is, the differential encoding success probability $P_i(\Delta_k)$ is identical for all states $i \in \mathbb{F}_{2^m}$, and can be denoted by $P(\Delta_k)$.

Corollary 11. *For a uniform Markov source, the limiting mean of sampled-scaled age $\tilde{A}$ is*

$$\mathbb{E}[\tilde{A}] = -\frac{1}{1-(p_a-p_d)P(\Delta_k)} + 1 + \frac{1}{1-p_a}.$$

and the distribution of peak age $\tilde{A}_M$ is given by

$$\Pr\{\tilde{A}_M = q\} = \begin{cases} s & q = 1, \\ (1-s)p_a^{q-2}(1-p_a) & q \geqslant 2. \end{cases}$$

where $s = (1-p_a) + (p_a - p_d)P(\Delta_k)$.

Proof: For uniform Markov source, the all ones column vector $\mathbf{1}$ is a right eigenvector for the block-diagonal matrix $\tilde{\mathbf{D}}$ corresponding to the eigenvalue $(p_a - p_d)P(\Delta_k)$. Hence, it follows that

$$\langle \nu(\mathbf{I} - \tilde{\mathbf{D}})^{-1}, \mathbf{1}\rangle = \frac{1}{1-(p_a-p_d)P(\Delta_k)}.$$

Result follows immediately from (7) and (10). ■

We note that setting $P(\Delta_k) = 1$ for the uniform Markov source gives us a *highly correlated source* wherein the consecutive state difference can always be represented by k bits. Such a source offers differential encoding opportunities after every successful update decoding, and hence serves as the lower bound on the age of the generalized update scheme for general Markov sources. A true update scheme ignores the differential encoding opportunity, as if $P(\Delta_k) = 0$, and serves as an upper bound on the age for general Markov source. Another special case is a uniform *iid* source which generates messages uniformly at random from $\mathbb{F}_{2^m}$, such that $P(\Delta_k) = 2^{k-m}$.

VIII. Numerical Comparison

Our proposed analysis is valid for any permutation invariant coding scheme. For clarity of exposition, we use a random coding scheme [33] for the numerical studies. For the random coding scheme, conditioned on the number of erasures E in an n-length codeword with $n-r$ parity bits, the probability of decoding failure [34] is given by

$$P(n, n-r, E) = 1 - \prod_{i=0}^{E-1}\left(1 - 2^{i-(n-r)}\right).$$

For all numerical comparisons, we have considered a Markov source $M(t)$ with the associated fundamental transition probability matrix $\mathbf{Q}$ to be tri-diagonal such that for all $i, j \in \mathbb{F}_{2^m}$, we have

$$\mathbf{Q}_{ij} \triangleq \Pr\{M(t+1) = j | M(t) = i\} = \begin{cases} 1-\alpha_i, & j = i, \\ \alpha_i/2, & |j-i| = 1. \end{cases}$$

The diagonal element $1-\alpha_i$ for each state i measures the self-correlation of source state i, with smaller α_i indicating higher self-correlation. For the numerical studies, we chose the number of information bits $m = 15$, and update codeword lengths to be $n = 20$ bits. Let Y be a standard normal random variable. For each $i \in \mathbb{F}_{2^m}$, the parameter α_i is chosen in $(0, 1)$ to create the tri-diagonal fundamental source state transition matrix $\mathbf{Q}$. Since the set of parameters $(\alpha_i : i \in \mathbb{F}_{2^m})$ is large (2^{15} elements), we chose these parameters randomly and independently from the following distribution for $x \in [0, 1]$,

$$\Pr\{X \leqslant x\} = P(|Y| \leqslant x || Y| \in [0, 1]) = \frac{\int_{-x}^{x} dy e^{-y^2/2}}{\int_{-1}^{1} dy e^{-y^2/2}}.$$

The above distribution corresponds to the absolute value of a standard normal random variable constrained between the duration [0, 1].

We plot the limiting average age of information of generalized incremental update scheme for the above mentioned source when the channel erasure probability $\epsilon \in \{0.01, 0.1, 0.2, 0.7\}$ in Fig. 6, as the differential encoding threshold k grows in $\{1, \dots, m\}$. As was alluded in the introduction, we observe the existence of optimal differential

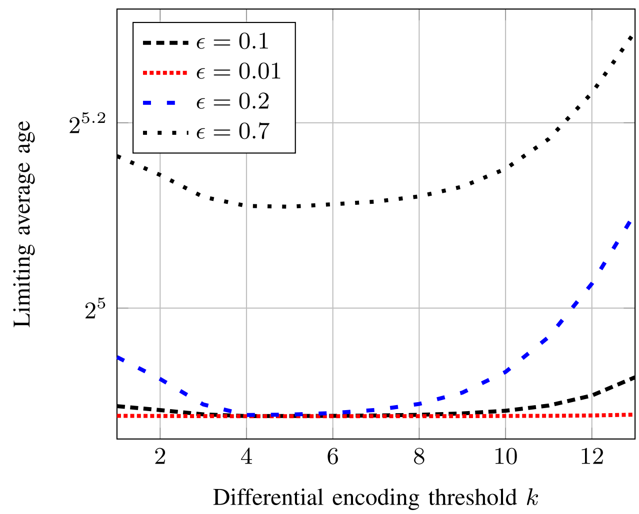

Fig. 6. We plot the variation of limiting average age with respect to the differential encoding threshold k in generalized incremental update scheme, when the channel erasure probability $\epsilon \in \{0.01, 0.1, 0.2, 0.7\}$. We have chosen codeword length $n = 20$ and the number of information bits $m = 15$.

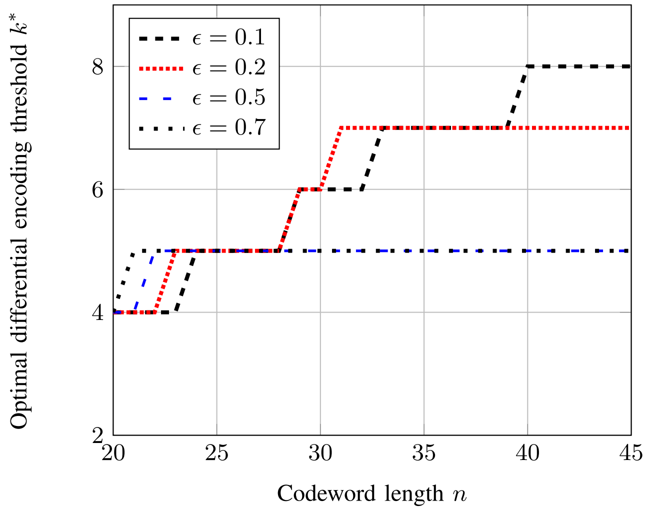

Fig. 7. We plot the variation of the optimal differential encoding threshold k^* for the generalized incremental update scheme, with respect to the codeword length n in $\{20, \ldots, 45\}$. The number of information bits is taken as $m = 15$, and the erasure probability is $\epsilon = \{0.1, 0.2, 0.5, 0.7\}$.

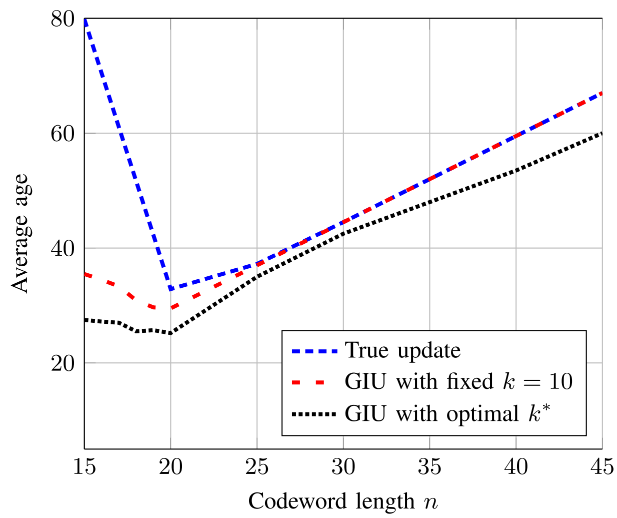

Fig. 8. We plot the limiting average age with respect to the codeword length n in $\{20, \ldots, 45\}$ for the generalized incremental update (GIU) with the optimal differential encoding threshold $k^*(n)$ and with a fixed differential encoding threshold k. We also plot the limiting average age with respect to codeword length n for true updates. We have chosen the number of information bits as $m = 15$, and the erasure probability as $\epsilon = 0.1$.

encoding threshold for generalized incremental updates of Markov sources.

In Fig. 7, we have plotted the age-optimal differential encoding level k^* of the generalized incremental update scheme for channel erasure probability $\epsilon \in \{0.1, 0.2, 0.5, 0.7\}$, as the codeword length n varies in $\{20, \ldots, 45\}$. We can infer from the plot that for a fixed value of ϵ, the optimal differential encoding level k^* first increases in code length and then saturates. The saturation of k^* results from the fact that as the codeword length n increases the source states $M(t+n)$ and $M(t)$ tend to get statistically independent, and hence there is very little correlation between these two source states to be exploited by the differential update. As the channel gets worse, the average age is dominated by decoding failures and hence adding more parities is more important than sending differential updates, and hence the saturation level for differential encoding threshold decreases with increase in channel erasure probability ϵ.

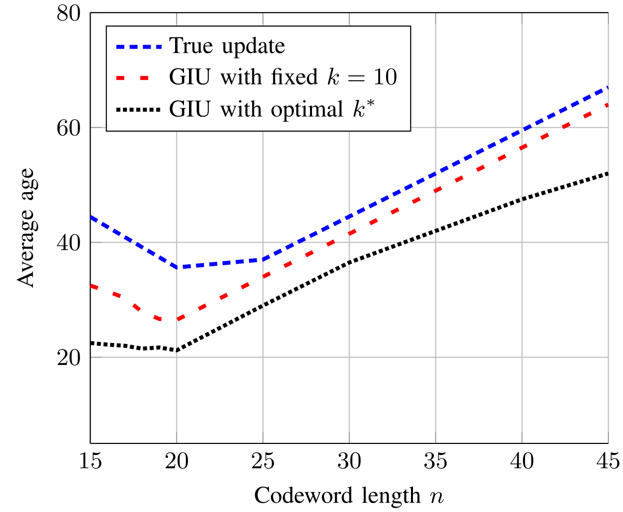

Fig. 9. This plot shows the variation of limiting average age with respect to the codeword length n in $\{20, \ldots, 45\}$ for the generalized incremental update (GIU) with the optimal differential encoding threshold $k^*(n)$ and with a fixed differential encoding threshold k. We also plot the limiting average age with respect to codeword length n for true updates. We have chosen the number of information bits as $m = 15$, and the erasure probability as $\epsilon = 0.01$.

In Fig. 8 and Fig. 9, we have plotted the limiting average age for two different channel erasure probabilities $\epsilon = 0.1$ and $\epsilon = 0.01$ respectively, as the code length n varies in the range $\{15, \ldots, 45\}$ for three different update schemes: the true update scheme, the generalized incremental update with fixed differential encoding threshold k, and the generalized incremental update with optimal differential encoding threshold $k^*(n)$. We note that for generalized incremental update, there exists an optimal differential encoding threshold $k^*(n)$ for each code length n that minimizes the average age. We have plotted the average age for parameters $(n, k^*(n))$ in the third case. We observe the existence of optimal code-length selection for age minimization for all three update schemes. Further, we observe that a sub-optimal selection of differential encoding threshold can significantly reduce the average age gain for the generalized incremental updates when compared to always sending true updates.

IX. Conclusion and Future Work

We considered a generalized incremental update scheme for real-time status updates for timely information reception at the receiver. This scheme exploits the temporal correlation between consecutive samples of a Markov source to reduce the information age at the receiver by sending differential updates, as compared to sending true updates. We numerically find the mean-age-optimal differential encoding threshold to decide between differential and actual updates. To this end, we needed to evaluate the age distribution of the proposed scheme for a general Markov source. We show that for this case, the source state and the information age are jointly Markov, and the standard technique of finding an age renewal interval can't be applied in a straightforward manner. Instead, utilizing the block structure of the corresponding transition matrix, we find the invariant distribution of the joint Markov process using matrix geometric methods. From this joint invariant distribution, we find the marginal stationary distribution of the age process. From the resulting limiting age distribution, we can compute limiting performance metrics such as mean age, peak age, decoding failure probability, and decay-rate of age-distribution for a general Markov source. These performance metrics admit a much simpler form for some special Markov sources with additional structure, and are derived as a corollary of the general result. We recover earlier results on highly correlated and uniform Markov sources, and provide new results for general Markov sources.

We showed that with the age-optimal differential encoding threshold, the generalized differential update scheme is more timely than the true update scheme. However, these gains seem to hinge on the accurate and immediate feedback from the receiver, and a perfect control message from the transmitter indicating the different encoding schemes for the differential and the true update. We remark that having immediate receiver feedback is not necessary for these gains to be realized. For example, consider the case when feedback delay is of one codeword duration. In this case, we can divide the source transmission into two streams that transmit n-length packets one after the other. Each stream receives its associated feedback before its next transmission opportunity. With this modification, both of these streams can employ the proposed update protocol, and can be studied in isolation using the proposed technique. However, the gains would diminish when the feedback messages are erased themselves, and if we need to send the control message along with the actual information over the unreliable channel. Consideration of unreliable feedback and control channels is an interesting future research direction.

While we consider the problem of age minimization for the periodic updates, it would be interesting to consider the problem of minimization of process estimation error at the receiver with non-periodic updates. For example, an interesting future direction is exploring sampling strategies outlined in [29] for a Markov source, with constraints on how often one can sample the source.

APPENDIX A
JOINT MARKOV PROCESS

We assume $A_0 = 1$ in the following.

Lemma 12. *The state space of random vector (M_j, A_j) is $\mathbb{F}_{2^m} \times \mathbb{N}$ for each $j \in \mathbb{N}$.*

Proof: The state space for the source state M_j is finite field $\mathbb{F}_{2^m}$. Since $A_0 = 1$, the sampled-scaled age $A_j \in \{1, A_{j-1}+1\}$ from the age evolution equation (4). Hence A_j can assume values in $\mathbb{N}$ for each j. ■

Lemma 13. *The discrete-time discrete-valued joint process $((M_j, A_j) \in \mathbb{F}_{2^m} \times \mathbb{N} : j \in \mathbb{N})$ is Markov.*

Proof: We denote the history of the joint sampled process until jth reception as $\mathcal{F}_j = \sigma(A_0, M_1, A_1, \dots, M_j, A_j)$. From the Markov property of the source and its independence from the erasure channel, it follows that for each $j \in \mathbb{N}$

$$P_{M_j|\mathcal{F}_{j-1}} = P_{M_j|M_{j-1}} = \mathbf{P}_{M_{j-1},M_j}. \tag{11}$$

We recall that the age at jth reception is $A_j = 1 + \bar{\xi}_j A_{j-1}$ from the age evolution equation (4). Further, the differential update indicator γ_j is a function of source states M_j, M_{j-1} and sampled-scaled age A_{j-1}, and can be written as

$$\gamma_j = \xi_{j-1} 1_{\{\delta_j \in \Delta_k\}} = 1_{\{A_{j-1}=1\}} 1_{\{M_j - M_{j-1} \in \Delta_k\}}.$$

Conditioned on the differential update indicator γ_j, the decoding success indicator ξ_j is independent of the past history $\mathcal{F}_{j-2}$. Hence, for each $j \in \mathbb{N}$

$$P_{A_j|\mathcal{F}_{j-1},M_j} = P_{A_j|M_j,M_{j-1},A_{j-1}}. \tag{12}$$

From the definition of conditional probability, and equations (11) and (12), we have

$$\begin{aligned} P_{M_j,A_j|\mathcal{F}_{j-1}} &= P_{A_j|\mathcal{F}_{j-1},M_j} P_{M_j|\mathcal{F}_{j-1}} \\ &= P_{A_j|M_j,M_{j-1},A_{j-1}} P_{M_j|M_{j-1}} = P_{M_j,A_j|M_{j-1},A_{j-1}}. \end{aligned} \tag{13}$$

Hence, we have obtained that, conditioned on the current state (M_j, A_j), the future state (M_{j+1}, A_{j+1}) is independent of the past history $\mathcal{F}_{j-1}$. This implies the Markov property for the joint process. ■

Lemma 14. *Conditioned on $A_{j-1} \geqslant 2$ and source states M_j, M_{j-1}, the probability distribution of sampled-scaled age A_j after jth reception, is given by*

$$P_{A_j|M_j,M_{j-1},A_{j-1}} = (1 - p_a) 1_{\{A_j=1\}} + p_a 1_{\{A_j=A_{j-1}+1\}}.$$

Conditioned on $A_{j-1} = 1$ and the source states M_j, M_{j-1}, the probability distribution of sampled-scaled age A_j after jth reception, is given by

$$P_{A_j|M_j,M_{j-1},A_{j-1}} = \begin{cases} 1 - p_a + (p_a - p_d) 1_{\{\delta_j \in \Delta_k\}}, & A_j = 1, \\ p_a - (p_a - p_d) 1_{\{\delta_j \in \Delta_k\}}, & A_j = 2. \end{cases}$$

Proof: From the age evolution equation (4), we have $A_j \in \{1, A_{j-1} + 1\}$ and $A_j = 1$ iff $\xi_j = 1$. Hence, we have

$$P_{A_j=A_{j-1}+1|M_j,M_{j-1},A_{j-1}} = \mathbb{E}[\bar{\xi}_j | M_j, M_{j-1}, A_{j-1}].$$

The event $\{A_{j-1} \geqslant 2\}$ is equivalent to the decoding failure of the $(j-1)$th update, i.e. $\xi_{j-1} = 0$. Hence the jth update has the true state information, and the corresponding differential update indicator $\gamma_j = 0$. From the conditional mean of decoding failure indicator (2), we have

$$\mathbb{E}[\bar{\xi}_j | M_j, M_{j-1}, A_{j-1} \geqslant 2] = \mathbb{E}[\bar{\xi}_j | \gamma_j = 0] = p_a.$$

Contrastingly, the indicator $1_{\{A_{j-1}=1\}}$ is equal to the decoding success indicator ξ_{j-1} of the $(j-1)$th update. Hence the jth update is differential update if $\{\delta_j \in \Delta_k\}$, and true update otherwise. From (3), it follows that

$$\mathbb{E}[\bar{\xi}_j | M_j, M_{j-1}, A_{j-1} = 1] = p_a 1_{\{\delta_j \notin \Delta_k\}} + p_d 1_{\{\delta_j \in \Delta_k\}}.$$

Result follows from the above two conditional means of the decoding failure indicator $\bar{\xi}_j$. ■

Lemma 15. *The joint process is a time-homogeneous Markov process.*

Proof: Homogeneity of the Markov process follows from the fact that the conditional distributions $P_{M_j|M_{j-1}}$ and

$P_{A_j|M_j,M_{j-1},A_j}$ depend only on the realized states, and don't vary with the update index. ■

Since the joint process is a time-homogeneous Markov chain, we can define the transition probability operator $\mathbf{T}$ for the joint process as

$$\mathbf{T}_{(M_{j-1},A_{j-1}),(M_j,A_j)} \triangleq P_{(M_j,A_j)|(M_{j-1},A_{j-1})}.$$

Corollary 16. *The probabilities for the joint process to transition from state (i,q) to state (i',q') for level $q \geqslant 2$ are*

$$\mathbf{T}_{(i,q),(i',q+1)} = p_a \mathbf{P}_{ii'}, \quad \mathbf{T}_{(i,q),(i',1)} = (1-p_a)\mathbf{P}_{ii'}.$$

The transition probabilities from level $q = 1$ are

$$\mathbf{T}_{(i,1),(i',2)} = (p_a - (p_a - p_d)1_{\{i'-i \in \Delta_k\}})\mathbf{P}_{ii'},$$
$$\mathbf{T}_{(i,1),(i',1)} = ((1-p_a) + (p_a - p_d)1_{\{i'-i \in \Delta_k\}})\mathbf{P}_{ii'}.$$

Proof: Recall that the sampled source process is Markov process with the transition matrix $\mathbf{P}$ independent of the age process. The result follows from equation (13) and Lemma 14. ■

Lemma 17. *The Markov process $((M_j, A_j) : j \in \mathbb{N})$ is irreducible.*

Proof: We fix two arbitrary states $i, l \in \mathbb{F}_{2^m}$ and level $q \geqslant 2$. To show irreducibility, it suffices to show that there is a positive probability of transition from state (l,q) to state $(i,1)$ in finite steps, and positive transition probability from state $(i,1)$ to (l,q).

We first show that there is a positive probability of transition from state (l,q) to state $(i,1)$ in finite steps. From the irreducibility of the Markov source, we can find $r \geqslant q-1$ such that $\mathbf{P}^r_{li} > 0$. The probability of r successful updates is lower bounded by $(1-p_a)^r$, and hence

$$\mathbf{T}^r_{(l,q),(i,1)} \geqslant \mathbf{P}^r_{li}(1-p_a)^r > 0.$$

Next, we consider the reverse transition from state $(i,1)$ to state (l,q). From the irreducibility of the Markov source, we can find $r \in \mathbb{N}$ such that $\mathbf{P}^r_{il} > 0$. One sample path to reach from state $(i,1)$ to (l,q) is to have $r-q+1$ decoding successes followed by $q-1$ consecutive failed updates. The probability of this sample path is lower bounded by $(1-p_a)^{r-q+1} p_d p_a^{q-2}$, and hence

$$\mathbf{T}^r_{(i,1),(l,q)} \geqslant \mathbf{P}^r_{il}(1-p_a)^{r-q+1} p_d p_a^{q-2} > 0.$$

This shows that the joint process is irreducible. ■

Lemma 18. *The Markov process $((M_j, A_j) : j \in \mathbb{N})$ is aperiodic.*

Proof: Since aperiodicity is a class property, we show that there exists a state i and some co-prime positive integers n_1, n_2 such that the probabilities of return to state $(i,1)$ in n_1 steps and in n_2 steps are positive. From aperiodicity of $\mathbf{P}$, it follows that there exists some state i and co-prime positive integers n_1, n_2 such that $\mathbf{P}^{n_1}_{ii} > 0$ and $\mathbf{P}^{n_2}_{ii} > 0$. If either n_1 or n_2 equals unity, then it follows that $\mathbf{P}_{ii} > 0$. Hence,

$$\mathbf{T}_{(i,1),(i,1)} = \mathbf{P}_{ii}(1-p_d) > 0.$$

Next, we consider $n_1, n_2 \geqslant 2$. Sample paths of return after $n_j - 1$ consecutive failures happen with non-zero probability greater than $(1-p_a)p_d p_a^{n_j-2}$ and hence for co-prime positive integers n_1, n_2 we have

$$\mathbf{T}^{n_j}_{(i,1),(i,1)} \geqslant \mathbf{P}^{n_j}_{ii}(1-p_a)p_d p_a^{n_j-2} > 0, \;\; j \in \{1,2\}.$$ ■

Lemma 19. *The joint process $((M_j, A_j) : j \in \mathbb{N})$ is positive recurrent.*

Proof: From our assumption of the irreducibility and aperiodicity of the source state transition matrix $\mathbf{P}$, it follows that the transition operator $\mathbf{T}$ for the joint process is also irreducible and aperiodic. Further, we know that the state space $\mathbb{F}_{2^m}$ of underlying sampled Markov source is finite and $A_j \in \mathbb{N}$. Therefore, the state space of the joint process is countably infinite, and it suffices to show that the mean of return time $T_{(i,1)}$ to a state $(i,1)$ is finite. Towards this, we construct a process $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$ with state space $\mathbb{F}_{2^m} \times \mathbb{N}$ for which the return time to state $(i,1)$ is stochastically larger than $T_{(i,1)}$, and the mean return time is finite.

At each decoding instant $j \in \mathbb{N}$, we set $\hat{M}_j = M_j$. While the decoding failure probability for the jth update is either p_a or p_d depending on (M_j, M_{j-1}, A_{j-1}), we set the decoding failure probability p_a for the constructed process $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$. Thus, we have

$$\Pr(\hat{A}_{j+1} = m | \hat{A}_j = q) = p_a 1_{\{m=q+1\}} + (1-p_a)1_{\{m=1\}}.$$

The coupled process $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$ is equivalent to a process which simply sends the true state at each transmission opportunity. Let $\hat{\xi}_j$ be the indicator for the decoding success for the jth received update for the constructed process, then we can couple the Bernoulli indicator random variables $\hat{\xi}_j$ and ξ_j such that $\hat{\xi}_j \leqslant \xi_j$ for each $j \in \mathbb{N}$ since $p_d < p_a$. That is, this coupling ensures that the success of an update (say the jth update) in the process $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$ implies the success of the corresponding update (the jth update) in the original process, $((M_j, A_j) : j \in \mathbb{N})$. This provides us a coupling of the constructed process $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$ to the original joint process $((M_j, A_j) : j \in \mathbb{N})$ such that $A_j \leqslant \hat{A}_j$ almost surely. For the coupled processes, we can see that the time to return to state $(i,1)$ for the process $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$ is stochastically larger than the corresponding time for the process $((M_j, A_j) : j \in \mathbb{N})$.

Next, we show that the constructed joint process has a unique stationary distribution which implies the mean return time to $(i,1)$ is finite for this process. Since the probability of decoding failures for the constructed process $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$ is always p_a, the sampled-scaled age process $(\hat{A}_j : j \in \mathbb{N})$ is independent of the source-state process $(\hat{M}_j : j \in \mathbb{N})$. This implies that for all $j \in \mathbb{N}$,

$$\Pr\{\hat{M}_j = m, \hat{A}_j = q\} = \Pr\{\hat{M}_j = m\}\Pr\{\hat{A}_j = q\}.$$

Hence, we can write the stationary distribution of the joint process as

$$\lim_{j\to\infty} \Pr\{\hat{M}_j = m, \hat{A}_j = q\}$$
$$= \lim_{j\to\infty} \Pr\{\hat{M}_j = m\} \lim_{j\to\infty} \Pr\{\hat{A}_j = q\},$$

whenever the limits on the individual terms exist. The stationary distribution for the source process $(\hat{M}_j : j \in \mathbb{N})$ is ν and for the process $\hat{A}_j$ we can compute the invariant distribution by using global balance equations, which is given by,

$$\lim_{j\to\infty} \Pr\{\hat{A}_j = q\} = (1 - p_a)p_a^{q-1}.$$

Thus the process $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$ has a unique invariant distribution given by

$$\lim_{j\to\infty} \Pr\{\hat{M}_j = m, \hat{A}_j = q\} = \nu_m(1 - p_a)p_a^{q-1}.$$

This implies the positive recurrence of $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$, and the result follows from the stochastic dominance of the mean return time to state $(i, 1)$ for the $((\hat{M}_j, \hat{A}_j) : j \in \mathbb{N})$ process over the corresponding return time for the joint process $((M_j, A_j) : j \in \mathbb{N})$. ■

References

[1] S. Poojary, S. Bhambay, and P. Parag, "Real-time status updates for correlated source," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2017, pp. 274–278.

[2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE Int. Conf. Comp. Commun. (INFOCOM)*, Mar. 2012, pp. 2731–2735.

[3] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Age of information with a packet deadline," in *Proc. IEEE ISIT*, Jul. 2016, pp. 2564–2568.

[4] K. Chen and L. Huang, "Age-of-information in the presence of error," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2579–2583.

[5] E. Najm and R. Nasser, "Age of information: The gamma awakening," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2574–2578.

[6] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 2666–2670.

[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," in *Proc. IEEE INFOCOM*, May 2016, pp. 1–9.

[8] P. Parag, A. Taghavi, and J.-F. Chamberland, "On real-time status updates over symbol erasure channels," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[9] R. Yates, E. Najm, E. Soljanin, and J. Zhong, "Timely updates over an erasure channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 316–320.

[10] S. Bhambay, S. Poojary, and P. Parag, "Fixed length differential encoding for real-time status updates," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2381–2392, Mar. 2019.

[11] A. Tridgell and P. Mackerras, "The rsync algorithm," Austral. Nat. Univ., Canberra, ACT, Australia, Tech. Rep. TR-CS-96-05, 1997.

[12] J. Mogul *et al.*, *Delta Encoding*, document RFC 3229 (Proposed Standard), Internet Engineering Task Force, Jan. 2002.

[13] J. D. Blischak, E. R. Davenport, and G. Wilson, "A quick introduction to version control with Git and GitHub," *PLoS Comput. Biol.*, vol. 12, no. 1, Jan. 2016, Art. no. e1004668.

[14] M. F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. New York, NY, USA: Marcel Dekker, 1989.

[15] G. Latouche and V. Ramaswami, "A logarithmic reduction algorithm for quasi-birth-death processes," *J. Appl. Probab.*, vol. 30, no. 3, pp. 650–674, Sep. 1993.

[16] B. Hajek, "Birth-and-death processes on the integers with phases and general boundaries," *J. Appl. Probab.*, vol. 19, no. 3, pp. 488–499, Sep. 1982.

[17] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "The stationary distribution of the age of information in FCFS single-server queues," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 571–575.

[18] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 1583–1587.

[19] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1897–1910, Apr. 2016.

[20] N. Pappas, J. Gunnarsson, L. Kratz, M. Kountouris, and V. Angelakis, "Age of information of multiple sources with queue management," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 5935–5940.

[21] Q. He, D. Yuan, and A. Ephremides, "On optimal link scheduling with min-max peak age of information in wireless systems," in *Proc. IEEE ICC*, May 2016, pp. 1–7.

[22] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2569–2573.

[23] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2018, pp. 1844–1852.

[24] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. IEEE ISIT*, Jun. 2017, pp. 576–580.

[25] E. Najm, R. Yates, and E. Soljanin, "Status updates through M/G/1/1 queues with HARQ," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 131–135.

[26] I. Kadota, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Minimizing the age of information in broadcast wireless networks," in *Proc. Allerton Conf.*, Sep. 2016, pp. 844–851.

[27] R. D. Yates and S. K. Kaul, "Status updates over unreliable multiaccess channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 331–335.

[28] M. Costa, S. Valentin, and A. Ephremides, "On the age of channel information for a finite-state Markov model," in *Proc. IEEE ICC*, Jun. 2015, pp. 4101–4106.

[29] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," in *Proc. IEEE ISIT*, Jun. 2017, pp. 321–325.

[30] J. P. Hespanha, "Modelling and analysis of stochastic hybrid systems," *IEE Proc.-Control Theory Appl.*, vol. 153, no. 5, pp. 520–535, Sep. 2006.

[31] A. R. Teel, A. Subbaraman, and A. Sferlazza, "Stability analysis for stochastic hybrid systems: A survey," *Automatica*, vol. 50, no. 10, pp. 2435–2456, 2014.

[32] P. Subrahmanya, R. Sundaresan, and D.-S. Shiu, "An overview of high-speed packet data transport in CDMA systems," *IETE Tech. Rev.*, vol. 21, no. 5, pp. 305–315, Sep. 2004.

[33] T. J. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY, USA: Cambridge Univ. Press, 2008.

[34] P. Parag, J. Chamberland, H. D. Pfister, and K. Narayanan, "Code-rate selection, queueing behavior, and the correlated erasure channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 397–407, Jan. 2013.

Sudheer Poojary (S'11–M'16) received the Ph.D. degree in 2017 from the Indian Institute of Science, Bengaluru, the M.Tech. degree in 2008 from Indian Institute of Technology, Bombay, and the B.E. degree in 2005 from Mumbai University. He is currently working in Qualcomm India Pvt. Ltd., Bengaluru as Senior Lead Engineer. Prior to that he was a post doctoral researcher at the University of Avignon, France. His research interests include communication systems, queueing theory, game theory and machine learning.

Sanidhay Bhambay (S'16) received the B.Tech. degree from JSS Noida, India, in 2014; and the M.Tech. (Research) degree from Indian Institute of Science, Bengaluru India, in 2018; all in electrical engineering. His research interests include interplay between queueing theory and error-control coding with applications in delay-sensitive communications.

He has been with Samsung Inc. Bengaluru, since 2018, working in the area of deep learning and machine learning.

Parimal Parag (S'04–M'11) is an assistant professor in the ECE department at Indian Institute of Science, Bengaluru. Prior to that, he was a senior system engineer (R&D) at Assia Inc. in Redwood City (2011-2014). He received a Ph.D. degree from the Texas A&M University in 2011, the M.Tech. and B.Tech. degrees in 2004 from IIT Madras, all in electrical engineering. He was a co-author of the 2018 IEEE ISIT student best paper, and a recipient of the 2017 early career award from the Science and Engineering Research Board.