2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)

Towards an "Effective Age" Concept

Clement Kam*, Sastry Kompella*, Gam D. Nguyen*, Jeffrey E. Wieselthier† and Anthony Ephremides‡

*Information Technology Division, Naval Research Laboratory, Washington, DC
†Wieselthier Research, Silver Spring, MD
‡Electrical and Computer Engineering Department, University of Maryland, College Park, MD

***Abstract*—Since for Markov signals age minimization generally implies prediction error minimization, we pursue in this paper the potential connection between age and what we call effective age. Effective age is loosely defined as an age-related metric that captures both the information structure of the signal and the sampling pattern that is used, and that it is minimized when the error is minimized. We consider several options for sampling mechanisms and signal models, and we evaluate age and prediction/estimation errors as steps in the quest for a meaningful effective age concept definition.**

I. Introduction

Until now, the studies of the age of information [1], [2] have focused mostly on calculating and minimizing the average age [3]–[5] or the peak age [6], [7]. We believe that minimizing age is useful because it implies minimizing prediction, or current value estimation error. If the signal that we monitor is a first-order Markov process, then it is clear that for an individual sample, the smaller the value of its age, the smaller the prediction or estimation error. However, when we consider *average* age (or average peak age), we know that the sampling pattern that minimizes them does not necessarily minimize the error. See for example the work of Yates et al. [8] on "just-in-time" sampling, or the work of Sun et al. [9] on the threshold policy optimality for minimizing age or error for a Wiener process.

The reason for this apparent disparity is that the value of the age is determined by two separate factors. One is the processing/transmission delay (including possible queueing delay). The other is the sampling pattern. Thus, a first concern is to consider modifying the "physical" or "actual" age so that its average value minimization (when we choose the sampling pattern) also achieves minimization of the prediction/estimation error. A second concern is to modify the definition of the physical age so as to capture more meaningfully the information content of the process. For example for a process with strong correlation properties, the physical age is not a good indicator of its prediction error. These concerns motivate our efforts to propose an "effective" age, the average value of which would correspond to the size of the error, and would capture the information structure (or content) of the process. This is of course a rather tall order, and in this paper we only begin to chart out a path towards these goals.

Our paper consists of two parts. In the first part, we outline some of our recent work on sampling patterns for the simplest of signals, namely a 2-state Markov Chain (generalized also to an N-state Markov Chain). We consider the "zero-wait" (also called "just-in-time") sampling pattern and compare the average age to that of a "threshold-like" sampling policy that generates a sample only when there is a state transition out of the current state. Our error parameter is the squared error in estimating the current value of the signal based on the most recent sample.

In the second part, we take a somewhat different approach where we focus on the micro-scale of the process values by considering "bit-by-bit" the encoded values of the process and consider a scheme of sampling (i.e., of what to transmit) that is akin to a threshold-like policy but focuses on a detailed view of the encoding properties. Our take on this approach is somewhat artificial in that it assumes that during each "clock"-time or slot, only one bit of the encoded symbol is transmitted. We believe that this restrictive model can be relaxed at some complexity cost, but it does offer some new insights.

Lastly, we present some effective age metrics that were introduced in our recent work and reassess whether they still satisfy our requirements in light of the current work. Our results are only tentative for the moment and rely mostly on simulation rather than analysis. We believe that the main value of our contribution consist of the key ideas behind our models and not so much the numerical results.

II. Threshold Sampling Policy

For a first-order Markov process, when estimating the current state based on a single sample, a lower age for the sample will yield a lower achievable estimation error. However, this does not necessarily extend to the average error when sampling and estimating a process in real time. It has been shown that in a remote monitoring system, in which there is a random delay from source to the monitor, for a Wiener process [9] and for a Markov process [10] the sampling strategy at the source that yields a lower average age does not necessarily yield a lower average error over the estimated process. In this work, we explore the relationship between age and error for a Markov process, and attempt to converge on a new concept of age, called "effective age."

A. Communication System

The communication system that we consider for sampling and prediction is a slotted-time system. If a sample is generated and thus available for transmission, the transmission delay S_i for the ith packet is an independent geometric random

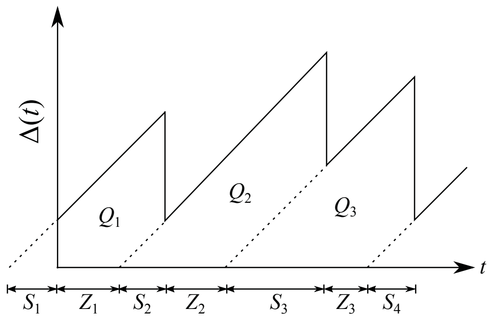


Fig. 1. Age function for "generate-at-will" model.

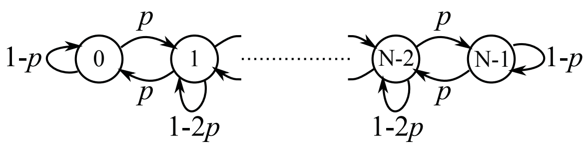


Fig. 2. N-state Markov chain.

variable with parameter $0 < \mu \leq 1$. The packet is fully received at the end of the transmission S_i. We assume the length of the time slots are equal to the transmission time for a Markov state. The sampling opportunities are just prior to the start of a new time slot, and there are no new samples generated until the transmission is successful. The source decides how many slots to wait, Z_i, before generating a new sample to begin transmission immediately. A sample path of the age function for such a system is shown in Fig. 1.

B. Markov Source

The signal that is being sampled is a discrete-time Markov source in which the time step is equal to a time slot length of the communication system. The Markov process is given by X_n, and the state transition is given by $\Pr(X_{n+1}|X_n)$. Although this model can be kept general, for the purpose of this work, we consider a special case in which we have an N-state birth-death type of Markov chain, where transitions only occur between state $i-1$ and i for $1 \leq i \leq N-1$, and self-transitions. Specifically, we consider the transition probabilities to be $\Pr(X_{n+1} = i|X_n = i-1) = \Pr(X_{n+1} = i-1|X_n = i) = p$, for $1 \leq i \leq N-1$. This Markov chain is shown in Figure 2.

C. Age vs. Prediction Error

The age of information of a monitored process is defined as $\Delta(t) = t - u(t)$, where $u(t)$ is the time of generation of the latest status update received at the destination. We are interested in the time-averaged age $\Delta = \lim_{T\to\infty} \int_0^T \Delta(t)\mathrm{d}t$.

At the destination, the monitor estimates (predicts) the current state at the source based on the states that it has received. For the Markov source described above, with $p < 0.5$, the estimate given the last received state $X_{u(t)}$ is chosen to be $\hat{X}_n = X_{u(t)}$. A reasonable error metric is the probability of error $e_{1,n} = \Pr(\hat{X}_n \neq X_n)$. Given the structure of the birth-death N-state Markov chain shown here, we can also use the squared error $e_{2,n} = (\hat{X}_n - X_n)^2$ as an error metric, where $X_n, \hat{X}_n \in \{0, 1, \ldots, N-1\}$.[1] The goal is to minimize the average estimation error, or in this case, the mean squared error (MSE) $\mathcal{E} = \lim_{N\to\infty}(1/N)\sum_{n=0}^{N} e_{2,n}$.[2] We are interested in the tradeoff between the average age and average MSE.

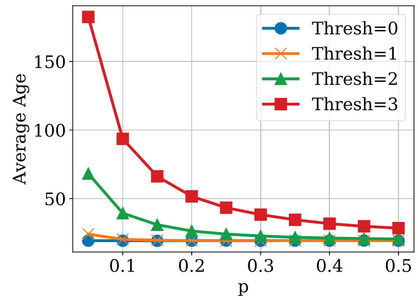


Fig. 3. Age vs. p, $\mu = 0.1$, 7 states.

D. Sampling Policy

The sampling policy that we study in this section is a discrete threshold policy, in which a sample is generated if there is no other sample currently being served and the distance between the last state sampled and the current state is greater than or equal to some threshold. The optimal threshold for the Wiener process (continuous state) was derived in [9]. The error probability was analyzed for the 2-state Markov process in [10] (called "zero-wait" and "sample-at-change" for threshold values of 0 and 1, respectively). It was shown that for this process, sampling at every opportunity (threshold of 0) yielded a smaller age but larger average error than sampling only when the state changed (threshold of 1). The average error probabilities were derived in [10] and are as follows:

$$\Pr(\mathcal{E}_{\mathcal{S}_0}) = \frac{p(2p + \mu(1-2p)(2-\mu))}{(\mu + 2p - 2\mu p)^2} \tag{1}$$

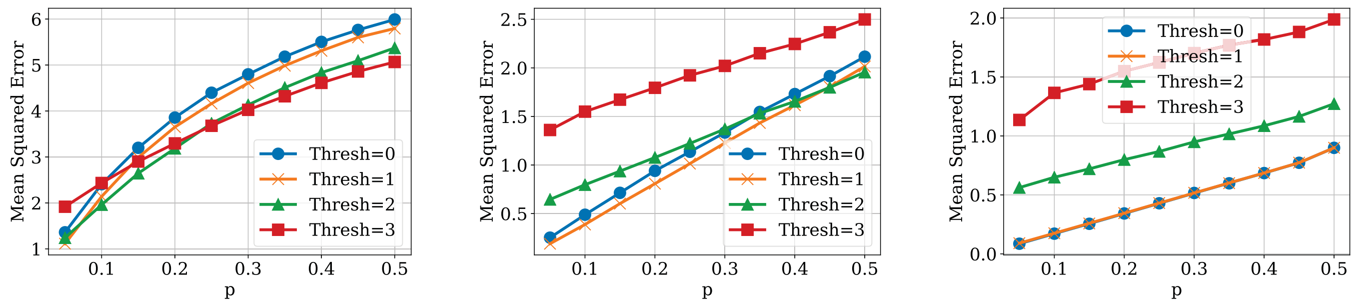
$$\Pr(\mathcal{E}_{\mathcal{S}_1}) = \frac{p(\mu + p - \mu p)}{\mu^2 + 2p(\mu + p - \mu(\mu + p))}. \tag{2}$$

E. Simulations

We simulated a 7-state Markov source to evaluate how the optimal threshold value changes with p. We ran each simulation so that there would be on average 200 state changes, and we averaged the age over time and over 1000 simulation runs. In Fig. 3, we have plotted the average age as a function of p for threshold values $0, 1, 2, 3$. As expected, the age is larger for larger thresholds, since the source waits for a larger change before sampling. In Fig. 4(a), we have plotted the MSE vs. p for the various threshold values. Clearly, the MSE does not follow the same trend or relative ordering between the threshold values as the age. We observe that for a threshold of 0, the MSE is less than when the threshold is 3 for $p \leq 0.1$,

[1]The squared error metric is dependent on the somewhat arbitrary labeling of states, since it is a discrete state space. For example, the states could take values $0, 1, 10, 100, \ldots$ instead of $0, 1, \ldots$. The labeling here is chosen to approximate a continuous state space.

[2]The estimator chosen is not the minimum mean squared estimator, so there is room for further improvement.



(a) $\mu = 0.1$ (b) $\mu = 0.5$ (c) $\mu = 1.0$

Fig. 4. MSE vs. p, 7 states.

but it has the highest MSE for $p \geq 0.15$. For a threshold of 1, the MSE is lowest for $p = 0.05$, but it increases rapidly with p relative to the cases where the threshold is 2 or 3. Setting the threshold to 2 yields a lower MSE for $p \leq 0.2$, but for $p \geq 0.25$, the MSE is lowest when the threshold is equal to 3. Thus, for a relatively low $\mu = 0.1$, a higher threshold than 1 should be used in most cases. This is probably due to the likelihood of a larger change between successful receptions, and that there is more value in transmitting a sample with a larger change when receptions occur less often.

We repeated the simulation with $\mu = 0.5$, and the age follows a similar trend to Fig. 3 but with lower age, so we omit it here. The MSE is plotted in Fig. 4(b), and we see that now a threshold of 3 yields the largest MSE, and a threshold of 2 is only best for $p \geq 0.45$. Applying similar reasoning from the $\mu = 0.1$ case, increasing the reception rate means that there are smaller changes between successful receptions, and receptions are successful enough that it need not wait for larger changes to improve the prediction. In most cases, a threshold of 1 provides the best prediction here.

We have also simulated for $\mu = 1.0$ and plotted the MSE in Fig. 4(c). In this case, using a threshold of zero or 1 yield the lowest MSE. The reasoning is the same as in the 2-state case, where transmission is successful on the first attempt, so there is no benefit in waiting to transmit. This is the case in which the relative ordering of the policies are virtually the same for the age as for the MSE.

III. Sampling at the Bit Level

In the second part of this paper, we consider the bit-level encoded values of the monitored process, and we study the sampling scheme at the time scale of bit transmissions. In the first part of this paper, the approach was to sample based on how much the state differed from the previous sample, whereas here we take the opposite approach and resample based on whether the state does *not* differ too much from the previous sample. We study the impact of what we call the *bit-level replacement* strategy on the age and prediction error, and we compare the performance to that of a threshold policy.

A. Communication System Model

For the communication system, we again have a time-slotted transmission as in Section II-A, but we assume an error-free channel with no random delay. However, instead of each time slot being equal to the time to transmit an entire state, each slot here is simply the time required to transmit one bit of information. Each state that is transmitted is represented by m bits. In general, we can substitute bits for symbols for non-binary signal constellations, but here we use bits for simplicity.

This model is clearly quite restrictive but it allows us to study the age at a timescale that is smaller than that of a single state transmission. This model is an error-free version of the model considered in [11].

B. Markov Source

Similar to the Markov source model in Section II-B, we have a discrete Markov process $X_0, X_1, \ldots$, but each time step is equal to the transmission time of a single bit, and each state is represented by m bits, in which X_n is one of 2^m discrete states, represented by m bits. Therefore, in this section we only consider Markov chains where the number of states is a power of 2. The transition probability from state i to state j is given by $\Pr(X_{n+1} = i | X_n = j)$.

C. Sampling Schemes

Since we have an error-free channel, a bit is successfully received in each slot, and we first consider sampling for continuous transmission schemes. The baseline scheme is zero-wait sampling and transmission of the *full state*, in which a new sample is taken after each m-bit state completes transmission, in every m time slots.

We also consider a *threshold* policy similar to that of Sec. II-D. If the threshold is not exceeded, no transmission occurs. If the threshold is exceeded, the source transmits the full state. It is possible to do differential encoding similar to [11], but it is beyond the scope of this work.

Finally, the continuous sampling scheme we focus on is called the *bit-level replacement* (BLR) scheme, in which we observe the state at each time slot, and if the bits of the partial state that have been sent match with the corresponding bits of the current state (mid-transmission of a full state), we send the remaining bits of the current state and refresh the time stamp with the current time. If there is no match between the current state and what has been sent, the remainder of the previous state completes transmission with the time stamp unchanged.

The more detailed example of how the BLR works is as follows:

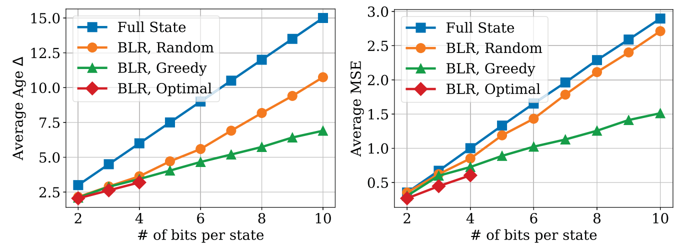
1) The transmission begins at time=0 by sending the first bit of the initial status (first bit of X_0).
2) At the next time slot (time=1), if the first bit of the current status matches that of the previous status, the second bit of the current status (X_1) is transmitted and the time stamp of the current status is updated to 1. However, if there is a mismatch in the first bit, the remaining bits of the initial status (time stamp = 0) are transmitted over the next $m-1$ slots, and the transmission process starts over again.
3) On the other hand, if there was a match in the first bit and the second bit of X_1 was transmitted, then in the next slot, the first two bits of the current status are compared to the bits that have already been transmitted. If the bits match, then the third bit of the current status (X_2) is transmitted and the time stamp of the status is updated to 2. However, if there is a mismatch in bits, the remaining bits of the previous status (X_1) are transmitted with time stamp = 1, and the transmission process starts over again.
4) This process continues with each remaining bit until there is a mismatch in the previously transmitted bits or until the status completes transmission, and the latest updated time stamp is provided.

D. Bit Encoding of States

The BLR allows for a newer state to preempt a previous state without having to start a new transmission, thus reducing the age. The performance of the BLR depends on the encoding of the bits to each state in the Markov chain, and this bit encoding can be optimized for BLR. The problem is to assign an m-bit codeword to each state.

1) Optimal algorithm: We devise a recursive algorithm that produces the optimal bit encoding. First, the states are divided into two partitions, and then each partition is likewise divided into two. This partitioning occurs a total of m times until each state is its own partition. At each step, the partitions are chosen such that the sum of the transition probabilities (weighted by the stationary probabilities of the starting states) between pairs of nodes in the same partition is maximized. The states of one partition are assigned a zero bit (in order of most significant bit) while those of the other partition are assigned the one bit. This repeats for every partition until all bits are assigned. Due to the complexity of having to consider all possible partitions and the sum of weighted transition probabilities, we are only able to compute the bit encoding for up to 4 bits (16 states).

2) Greedy algorithm: To handle larger numbers of states, we devised a suboptimal greedy algorithm for the bit encoding. First, the weighted transition probabilities are sorted from largest to smallest. The starting and ending states for the largest weighted transition probability are assigned to any two codewords that have the most consecutive significant bits in common. Those codewords are removed from the set of available m-bit codewords for assignment. For each successive



(a) Age vs. # bits per state (b) MSE vs. # bits per state

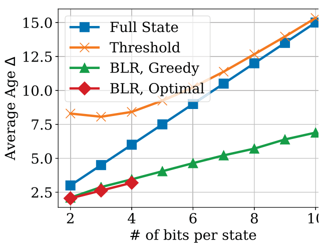
Fig. 5. Full State vs. BLR, p=0.1

transition probability, it is possible that both, one, or neither of the starting or ending states has already been assigned a codeword. If both states have been assigned a codeword, then the algorithm simply moves on to the next transition probability. If exactly one of the states is already assigned a codeword, the other state is assigned the codeword that has the most consecutive significant bits in common with the already assigned state. If neither state has been assigned a codeword, then the they are assigned the two available codewords with the most consecutive significant bits in common. This algorithm iterates through each transition probability from largest to smallest until all states have been assigned a codeword.

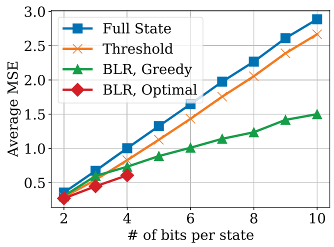
E. Simulations

We simulated the bit-by-bit communication system for the various sampling strategies. We start with a comparison between continuous transmission schemes with and without bit-level replacement. In Figure 5, we plot the age and the MSE vs. the number of bits per state m, so that the number of states is increasing by a factor of 2. The value of p for the Markov chain is set to 0.1. We observe that the BLR schemes outperform the full state scheme in terms of age and MSE, and that the relative ordering is preserved from the age to MSE. For this error-free, deterministic delay system, the relative ordering to be preserved, as we observed in Sec. II-E, such that lower age implies lower prediction error. The optimal bit encoding does the best, but we are unable to generate results beyond 4 bits per state due to the computational complexity of the recursive algorithm.

Next we compare the bit-level replacement to threshold-based sampling. Although not shown here, we observed that for this communication system, a threshold of 1 was better than a threshold of 2 in terms of both age and MSE. In Figure 6, we compare the threshold scheme (threshold=1) with the full state scheme (threshold=0) and the Greedy BLR scheme for the case where $p = 0.1$. We observe that the age for the threshold policy is the worst (Fig. 6(a)), but unlike in Figure 4(c), the error performance is better than that of the full state scheme. This is because the states can change at any of the m bit transmission of a full state, so delaying until the state changes can improve the MSE. However, it is not as good as the Greedy BLR scheme for $m \geq 4$ (Fig. 6(b)). The Optimal BLR scheme is the best in both age and MSE wherever it can be evaluated.



(a) Age vs. # of bits per state



(b) MSE vs. # of bits per state

Fig. 6. Full State vs. BLR vs. Threshold=1, $p = 0.1$

IV. Effective Age Metrics

In [10], we proposed two effective age metrics that are minimized when the error is minimized. The first is the *sampling age*, which conveys the age relative to some ideal sampling pattern. The idea is that in the 2-state Markov source system, the age that matters with regards to estimation error is not the actual AoI, but rather the age of the sampling instant relative to the time the status changed (and should have been sampled). Assuming we have a sampling pattern that optimizes the error, we define $g(t)$ to be the most recent "optimal" sampling time relative to time t. We then define $s(t)$ to be the first actual sampling time following $g(t)$. If no sample has occurred between $g(t)$ and t, we let $s(t) = t$. Finally, the sampling age is defined as $\Delta_{samp}(t) = s(t) - g(t)$. If the actual sampling pattern coincides with the optimal sampling pattern, or even if there are samples in between the optimal samples, the sampling age is zero for all t. Otherwise, the sampling age increases linearly after each optimal sampling instant until the actual sampling instant, at which time it remains constant until the next optimal sampling instant.

This age metric may work for the first communication model studied here, but it needs to be amended for other models, such as queueing systems, to account for the impact of oversampling. Specifically, we define $s(t)$ not as the first time a sample is taken and placed in a queue, but we define it as the time a sample (taken after $g(t)$) first enters service. Furthermore, this metric does not directly apply to the bit-by-bit communication model, which re-samples before completing transmission of a state, so a metric should be tailored to such a model.

The second metric proposed is the *cumulative marginal error* (CME), defined as $\Delta_{CME}(t) = \int_{r(t)}^{t} h(\tau)\mathrm{d}\tau$, where $h(t)$ is some penalty function, and $r(t)$ is the reception time of the packet received with most recent timestamp. The idea here is that for each time period between samples, we accumulate the total error penalty (or estimated error) as time progresses, since we are interested in prediction error. For MSE, we propose letting $h(\tau)$ be the squared error at time τ. Further investigation is needed into whether this choice yields a relative ordering of the CME that is the same as for the MSE.

V. Conclusion

In this work, we investigated the impact of the sampling policy on the age of information and the error in predicting the current state of a Markov source. We consider two variations on the threshold-based sampling idea. First we consider sampling only after the distance between the current state and the last sampled state meets a threshold. For longer service time, a larger threshold tends to do better, particularly for higher probability of transition in the Markov chain. In the second case, we consider sampling and transmission at the bit level. We propose resampling while we are still transmitting a state, provided that the current state has not veered too far from the state(s) already sampled during a single state transmission, a policy we call bit-level replacement (BLR). The performance depends on the bit assignment of the states, and we provide an optimal and greedy bit assignment algorithm. Our simulations demonstrate that greedy BLR outperforms the zero-wait and sample-at-change (threshold of 0 and 1, respectively) policies in both age and prediction error in most cases. Finally we reintroduce our proposed effective age metrics and make some new observations. Further study into these and possibly other metrics is needed in light of the current work.

Acknowledgment

This work was supported by the Office of Naval Research.

References

[1] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, June 2011, pp. 350–358.

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

[3] ——, "Status updates through queues," in *Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2012, pp. 1–6.

[4] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, Jul. 2012, pp. 2666–2670.

[5] C. Kam, S. Kompella, and A. Ephremides, "Age of information under random updates," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 66–70.

[6] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1897–1910, April 2016.

[7] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1681–1685.

[8] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 3008–3012.

[9] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the wiener process over a channel with random delay," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 321–325.

[10] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Towards an effective age of information: Remote estimation of a Markov source," to appear in 2018 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Apr. 2018.

[11] S. Bhambay, S. Poojary, and P. Parag, "Differential encoding for real-time status updates," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2017, pp. 1–6.