Optimal Lossless Source Codes for Timely Updates

Prathamesh Mayekar[†]   Parimal Parag[†]   Himanshu Tyagi[†]

Abstract—A transmitter observing a sequence of independent and identically distributed random variables seeks to keep a receiver updated about its latest observations. The receiver need not be apprised about each symbol seen by the transmitter, but needs to output a symbol at each time instant t. If at time t the receiver outputs the symbol seen by the transmitter at time $U(t) \leq t$, the age of information at the receiver at time t is $t - U(t)$. We study the design of lossless source codes that enable transmission with minimum average age at the receiver. We show that the asymptotic minimum average age can be attained (up to a constant bits gap) by Shannon codes for a tilted version of the original pmf generating the symbols, which can be computed easily by solving an optimization problem. Underlying our construction for minimum average age codes is a new variational formula for integer moments of random variables, which may be of independent interest.

I. INTRODUCTION

Timeliness is emerging as an important requirement for communication in cyber-physical systems (CPS). Broadly, it refers to the requirement of having the latest information from the transmitter available at the receiver in a timely fashion. It is important to distinguish the requirement of timeliness from that of low delay transmission: The latter places a constraint on the delay in transmission of each message, while timeliness is concerned about how recent is the current information at the receiver. In particular, even if a message m is transmitted with low delay, if the receiver has to wait for subsequent messages, the information conveyed by message m looses its timeliness. A heuristically appealing metric that captures timeliness of information, termed its *age*, was proposed in [1] (see [2]–[6] for subsequent developments) for a setting involving queueing and link layer delays. In this paper, we initiate a

systematic study of the design of lossless source codes with the goal of minimizing the age of the information at the receiver.

Specifically, we consider the problem of source coding where a transmitter receives symbols generated from a known distribution and seeks to communicate them to a receiver in a timely fashion. To that end, it encodes each symbol x to $e(x)$ using a variable length prefix-free code e. The coded sequence is then transmitted over a noiseless communication channel that sends one bit per unit time. We restrict our treatment to a simple class of deterministic[1] update schemes, termed *memoryless update schemes*, where the transmitter cannot

[†]Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India. Email: {prathamesh, parimal, htyagi}@iisc.ac.in

[1]Our analysis of average age extends to randomized schemes as well; see Section VI.
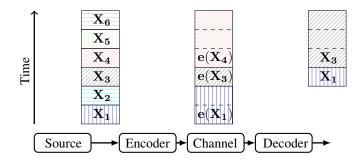

Fig. 1: An example of the proposed memoryless update scheme, where the symbol X_2 is skipped since the channel was busy sending X_1 when it arrived. Furthermore, the decoder retains output X_3, a symbol which was received within a delay of 1 time unit, for 3 time slots while X_4 was being sent.

store the symbols it has seen previously and sends the next observed symbol once the channel is free.

On the receiver side, at each instance t the decoder outputs a time $U(t)$ and the symbol $X_{U(t)}$ seen by the transmitter at time $U(t)$. Thus, the age of information at the receiver at time t is given by $A(t) = t - U(t)$. We illustrate the setup in Figure 1.

Our goal in this paper is to design prefix-free codes for which the average age of the memoryless scheme above is minimized, namely codes e which minimize

$$\bar{A}(e) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} A(t).$$

This formulation is apt for the timely update problem where the transmitter need not send each update and strives only to reduce the average age of the information at the receiver.

Using the renewal reward theorem, we derive a closed form formula for the asymptotic average age attained by a prefix-free code. Interestingly, this formula is a rational function of the first and the second moment of the random codeword length. Our main contribution in this paper is a variational formula for the second moment of random variables that enables an algorithm for finding the code that attains the minimum asymptotic average age up to a constant gap. The variational formula is of independent interest and may be useful in other settings where such cost functions arise; we point-out one such setting in the final section. In fact, our prescribed prefix-free code is a Shannon code for a tilted version of the original pmf which can be computed by solving a simple optimization problem.

The aforementioned formula for average age implies an $O(\log|\mathcal{X}|)$ upper bound on the minimum average age, attained by a fixed length code. We show that the same upper bound of $O(\log|\mathcal{X}|)$ holds for the average age of a Shannon code for the original distribution as well. However, we exhibit an example where Shannon codes for the original distribution have $\Omega(\log|\mathcal{X}|)$ age, while our proposed scheme yields an average age of $O(\sqrt{\log|\mathcal{X}|})$.

Note that the problem of designing update codes with low average age is related to real-time source coding ($cf.$ [7]) where we seek to transmit a stream of data under strict delay bounds. A related formulation has also emerged in the control over communication network literature ($cf.$ [8]) where an observation is quantized and sent to an estimator/controller to enable control. Here, too, the requirement is that of communication under bounded delay. Our proposed minimum average age problem differs from both these formulations since we need not send the entire stream and are allowed to skip some symbols. In our applications of interest, the allowed communication rates are much lower than the rate at which data is generated, and there is no hope of transmitting all the data at bounded delay, as mandated by the formulations available hitherto.

The next section contains a formal description of our setting and a formula for asymptotic average age of a code. Our main technical tool is presented in Section III, and we apply it to the minimum average age code design problem in Section IV. Numerical evaluations of our proposed scheme for Zipf distribution is presented in Section V. We conclude with discussion on extensions and the minimum queuing delay problem in the final section.

II. AVERAGE AGE FOR MEMORYLESS UPDATE SCHEMES

Consider a system in which at every time instant t, a transmitter observes a symbol X_t generated from a finite alphabet $\mathcal{X}$ with pmf P. We assume that the sequence $\{X_t\}_{t=1}^{\infty}$ is independent and identically distributed (iid). The transmitter has a noiseless communication channel at its disposal over which it can transmit one bit per unit time. A *memoryless update scheme* consists of a prefix-free code, represented by its encoder $e : \mathcal{X} \to \{0,1\}^*$, and a decoder which at each time instant t declares a time index $U(t) \le t$ and an estimate $\hat{X}_{U(t)}$ for the symbol $X_{U(t)}$ that was observed by the encoder at time $U(t)$. We focus on error-free schemes and require $\hat{X}_{U(t)}$ to equal $X_{U(t)}$ with probability 1.

In a memoryless update scheme, once the encoder starts communicating a symbol x, encoded as $e(x)$, it only picks up the next symbol once all the bits in $e(x)$ have been transmitted successfully to the receiver. The time index $U(t)$ is updated to a new value only upon receiving all the encoded bits for the current symbol. That is, if the transmission of a symbol is completed at time $t-1$, the encoder will start transmitting X_t in the next instant. Morever, if the final bit of $e(X_t)$ is received at time t', $U(t')$ is updated to t; else it remains unchanged. A typical sample path for $U(t)$ is given in Figure 2. The age
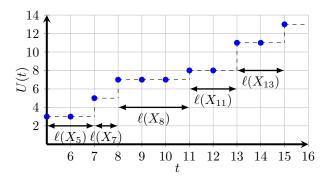

Fig. 2: A sample path of $U(t)$ with $U(5) = 3$.

$A(t)$ of the symbol available at the receiver at time t is given by

$$A(t) = t - U(t).$$

Note that it is natural to allow errors in estimates of $X_{U(t)}$ as well as allow encoders with memory, but we limit ourselves to the simple error-free and memoryless setting in this paper.

We are interested in designing prefix-free codes e that minimize the average age for the memoryless update scheme described above.

Definition II.1. The *average age* for a prefix-free code e, denoted $\bar{A}(e)$, is given by

$$\bar{A}(e) = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} (t - U(t)).$$

While $\bar{A}(e)$ is a random variable, we will prove that it is a constant almost surely. Let $\ell(x)$ denote the length of the codeword $e(x)$, $x \in \mathcal{X}$, and $L = \ell(X)$ with $X \sim P$ denote the random code-length. The result below uses the renewal reward theorem to provide a closed form expression for $\bar{A}(e)$ in terms of the first and the second moments of L.

Theorem II.2. *Consider a random variable X with pmf P on $\mathcal{X}$. For a prefix-free code e, the average age $\bar{A}(e)$ is given by*

$$\bar{A}(e) = \mathbb{E}[L] + \frac{\mathbb{E}[L^2]}{2\mathbb{E}[L]} - \frac{1}{2} \quad a.s.. \tag{1}$$

Denoting by $\bar{A}^*$ the minimum average age over all prefix-free codes e, as a corollary of the characterization above, we can obtain the following bounds for $\bar{A}^*$.

Corollary II.3. *For any pmf P over $\mathcal{X}$, the optimal average age $\bar{A}^*$ is bounded as*

$$\frac{3}{2}H(P) - \frac{1}{2} \le \bar{A}^* \le \frac{3}{2}\log|\mathcal{X}| + 1.$$

The proof of lower bound simply uses Jensen's inequality and the fact that $\mathbb{E}[L] \ge H(P)$ for a prefix free code; the upper bound is obtained by using codewords of constant length $\lceil \log|\mathcal{X}| \rceil$.

Note that the lengths $\ell(x)$ are required to be nonnegative integers. However, for any set of real-valued lengths $\ell(x) \ge 0$,

we can obtain integer-valued lengths by using the rounded-off values $\lceil \ell(x) \rceil$. Unlike the average length cost, the average age cost function identified in (1) is not an increasing function of the lengths. Nevertheless, by (1), the average age $\bar{A}(e)$ achieved when we use the rounded-off values is no more than

$$\mathbb{E}[L] + \frac{\mathbb{E}[L^2]}{2\mathbb{E}[L]} + 2. \tag{2}$$

Accordingly, in our treatment below we shall ignore the integer constraints and allow nonnegative real valued length assignments.

Returning now to the bound of Corollary II.3, the upper and lower bounds coincide when P is uniform. Recall that a Shannon code for P assigns lengths $\ell_S(x) = \lceil -\log P(x) \rceil$ to a symbol x. In view of the foregoing discussion, Shannon codes for a uniform distribution attain the minimum average age. The next result gives an upper bound on average age for Shannon codes for an arbitrary P on $\mathcal{X}$.

Lemma II.4. *Given a pmf P on $\mathcal{X}$, a Shannon code e for P has average age at most $O(\log |\mathcal{X}|)$.*

The proof is based on noting that for the distribution P' with $P'(x) \propto \ell_S(x)P(x)$, we have $H(P') \leq \log |\mathcal{X}|$, which in turn implies that $\mathbb{E}[L^2]/\mathbb{E}[L]$ is $O(\log |\mathcal{X}|)$ when $L = \ell_S(X)$.

It is of interest to examine if, in general, a Shannon code for P itself has average age close to $\bar{A}^*$, as was the case for the uniform distribution. In fact, it is not the case. Below we exhibit a pmf P where the average age of a Shannon code for P is $\Omega(\log |\mathcal{X}|)$, yet a Shannon code for another distribution (when evaluated for P) has an average age of only $O(\sqrt{\log |\mathcal{X}|})$.

Example II.5. Consider $\mathcal{X} = \{0, ..., 2^n\}$ and a pmf P on $\mathcal{X}$ given by

$$P(x) = \begin{cases} 1 - \frac{1}{n}, & x = 0 \\ \frac{1}{n2^n}, & x \in \{1, \ldots, 2^n\}. \end{cases}$$

Using (1), the average age $\bar{A}(e_P)$ for a Shannon code for P can be seen to satisfy $\bar{A}(e_P) \approx (n + 2\log n)/4$. On the other hand, if we instead use a Shannon code for the pmf P' given by

$$P'(x) = \begin{cases} \frac{1}{2\sqrt{n}}, & x = 0 \\ \frac{1 - 2^{-\sqrt{n}}}{2^n}, & x \in \{1, \ldots, 2^n\}, \end{cases}$$

we get $\mathbb{E}[L] \approx \sqrt{n}$ and $EL^2 \approx 2n$, whereby $\bar{A}(e_{P'}) \approx 2\sqrt{n}$, just $O(\sqrt{\log |\mathcal{X}|})$. □

Thus, one needs to look beyond the standard Shannon codes for P to find codes with minimum average age. Interestingly, we show that Shannon codes for a tilted version of P attain the optimal asymptotic average age (up to the constant loss of atmost 2.5 bits incurred by rounding-off lengths to integers). In particular, for the example above, our proposed optimal codes will have an average age of only $O(\sqrt{\log |\mathcal{X}|})$ in comparison to $O(\log |\mathcal{X}|)$ of Shannon codes for P.

A key technical tool in design of our codes is a variational formula that will allow us to linearize the cost function in (1), thereby rendering Shannon codes for a tilted distribution optimal. We present this in the next section.

III. A VARIATIONAL FORMULA FOR p-NORM

The expression for average age identified in Theorem II.2 involves the second moment of the random codeword length L. This is in contrast to the traditional variable length source coding problem where the goal is to minimize the average codeword length $\mathbb{E}[L]$. For this standard cost, Shannon codes which assign a codeword of length $\lceil -\log P(x) \rceil$ to the symbol x come within 1-bit of the optimal cost (see, for instance, [9]). A variant of this standard problem was studied in [10], where the goal was to minimize the log-moment generating function $\log \mathbb{E}[\exp(\lambda L)]$. A different approach for solving this problem is given in [11] where the *Gibbs variational principle* is used to linearize the nonlinear cost function $\log \mathbb{E}[\exp(\lambda L)]$. The next result provides the necessary variational formula to extend the aforementioned approach to another nonlinear function, namely $\|L\|_p := (\mathbb{E}[L^p])^{\frac{1}{p}}$ for $p > 1$.

Theorem III.1. *For a random variable X with distribution P and $p \geq 1$ such that $\|X\|_p < \infty$, we have*

$$\|X\|_p = \max_{Q \ll P} \mathbb{E}\left[\left(\frac{dQ}{dP} \right)^{\frac{1}{p'}} |X| \right],$$

where $p' = p/(p-1)$ is the Hölder conjugate of p.

Proof. For $Q \ll P$ and $0 < \alpha \neq 1$, let $D_\alpha(P, Q)$ denote the Rényi divergence of order α between distributions Q and P (see [12]), given by

$$D_\alpha(P, Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_P\left[\left(\frac{dQ}{dP} \right)^\alpha \right].$$

It is well-known that $D_\alpha(P, Q) \geq 0$ with equality iff $P = Q$. Consider the probability measure $P_p \ll P$ be given by

$$\frac{dP_p}{dP} = \frac{1}{\|X\|_p^p} \cdot |X|^p.$$

Then, for $\alpha = 1/p'$,

$$0 \leq D_\alpha(P_p, Q) = \frac{1}{\alpha - 1} \log \mathbb{E}\left[\left(\frac{dQ}{dP} \right)^\alpha \left(\frac{dP_p}{dP} \right)^{1-\alpha} \right]$$

$$= -p \log \mathbb{E}\left[\left(\frac{dQ}{dP} \right)^\alpha |X| \right] + p \log \|X\|_p,$$

where the previous equality holds since $p(1 - \alpha) = 1$. Thus, for every $Q \ll P$,

$$\mathbb{E}\left[\left(\frac{dQ}{dP} \right)^\alpha |X| \right] \leq \|X\|_p,$$

with equality iff $P_p = Q$. □

IV. PREFIX-FREE CODES WITH MINIMUM AVERAGE AGE

We now present a recipe for designing prefix-free codes with minimum average age. By Theorem II.2, we seek prefix-free codes that minimize the cost

$$\mathbb{E}[L] + \frac{\mathbb{E}[L^2]}{2\mathbb{E}[L]}, \tag{3}$$

where $L = \ell(X)$ for X with pmf P. As is well known, a prefix-free code with lengths $\{\ell(x) \in \mathbb{N}, x \in \mathcal{X}\}$ exists iff lengths satisfy Kraft's inequality ($cf.$ [9]), $i.e.$, iff

$$\sum_{x \in \mathcal{X}} 2^{-\ell(x)} \leq 1. \tag{4}$$

Following the discussion leading to (2), we relax the integral constraints for $\ell(x)$ and search over all real-valued $\ell(x) \geq 0$ satisfying (4). Specifically, we search for optimal lengths over the set $\Lambda = \{\ell \in \mathbb{R}_+^{|\mathcal{X}|} : \sum_{x \in \mathcal{X}} 2^{-\ell(x)} \leq 1\}$. As noticed in (2), this can incur a loss of only a constant. A key challenge in minimizing (3) is that it is nonlinear. We linearize this cost by relying on Theorem III.1 as follows:

1) Note that

$$\mathbb{E}[L] + \frac{\mathbb{E}[L^2]}{2\mathbb{E}[L]} = \max_{z \geq 0}\left(1 - \frac{z^2}{2}\right)\mathbb{E}[L] + z\|L\|_2.$$

2) Then, Theorem III.1 yields $\mathbb{E}[L] + \frac{\mathbb{E}[L^2]}{2\mathbb{E}[L]}$ equal to

$$= \max_{z \geq 0}\left(1 - \frac{z^2}{2}\right)\mathbb{E}[L] + z \max_{Q \ll P}\sum_{x \in \mathcal{X}}\sqrt{Q(x)P(x)}\ell(x),$$

$$= \max_{z \geq 0}\max_{Q \ll P}\sum_{x \in \mathcal{X}} g_{z,Q,P}(x)\ell(x),$$

where

$$g_{z,Q,P}(x) := \left(1 - \frac{z^2}{2}\right)P(x) + z\sqrt{Q(x)P(x)}. \tag{5}$$

Thus, our goal is to identify the minimizer ℓ^* that achieves

$$\Delta^*(P) = \min_{\ell \in \Lambda}\max_{z \geq 0}\max_{Q \ll P}\sum_{x \in \mathcal{X}} g_{z,Q,P}(x)\ell(x). \tag{6}$$

The result below captures our main observation and facilitates the computation of optimal lengths attaining $\Delta^*(P)$.

Theorem IV.1. *The optimal minimax cost $\Delta^*(P)$ in* (6) *satisfies*

$$\Delta^*(P) = \max_{z \geq 0}\max_{Q \ll P}\min_{L \in \Lambda}\sum_{x \in \mathcal{X}} g_{z,Q,P}(x)\ell(x)$$

$$= \max_{\substack{z \in \mathbb{R}, Q \ll P, \\ (z,Q) \in \mathcal{G}}}\sum_{x \in \mathcal{X}} g_{z,Q,P}(x)\log\frac{\sum_{x' \in \mathcal{X}} g_{z,Q,P}(x')}{g_{z,Q,P}(x)}, \tag{7}$$

where $\mathcal{G} := \{z \in \mathbb{R}, Q \in \mathbb{R}^{|\mathcal{X}|} : g_{z,Q,P}(x) \geq 0 \quad \forall x \in \mathcal{X}\}$. Furthermore, if (z^*, Q^*) is the maximizer of the right-side of (7)*, then the minimax cost* (6) *is achieved by an optimal average length code for the pmf P^* on $\mathcal{X}$ given by*

$$P^*(x) = \frac{g_{z^*,Q^*,P}(x)}{\sum_{x' \in \mathcal{X}} g_{z^*,Q^*,P}(x')}.$$

Thus, our prescription for design of update codes is simple: Use a Shannon code for P^* instead of P. To compute P^*, we need to solve the optimization problem in (7). This problem is concave in Q for each fixed z and is concave in z for each fixed Q, but may not be jointly concave in (z, Q). Nevertheless, we can solve it using standard numerical packages; see next section for further discussion. Also, it is intriguing to examine how much our recipe gains over a Shannon code for the original distribution P. This, too, will be discussed in the next section.

V. NUMERICAL RESULTS FOR ZIPF DISTRIBUTION

We now illustrate our recipe for construction of prefix-free codes that yield minimum average age for memoryless update schemes when P is a Zipf distribution[2]. Specifically, we illustrate our qualitative results using the `Zipf`(s, N) distribution with alphabet $\mathcal{X} = \{1, ..., N\}$ and given by $P(i) = \frac{i^{-s}}{\sum_{j=1}^{N} j^{-s}}$, $1 \leq i \leq N$. When the parameter s is close to 0, the `Zipf`(s, N) distribution approaches a uniform distribution, and therefore, as seen earlier, Shannon codes for P are close to optimal. However, for larger values of s, we note in Figure 3 that our recipe yields prefix-free codes with smaller average age than Shannon codes for P. When we round-off real lengths to integers, the gains are subsided but still exist. The distribution P^* we use to construct our codes
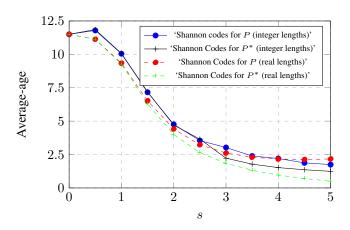

Fig. 3: Comparison of proposed codes and Shannon codes for `Zipf`$(s, 256)$ with varying s. The average age is computed using real-valued lengths as well as lengths rounded-off to integer values.

seems to be a flattened version of the original Zipf distribution; we illustrate the two distributions for `Zipf`$(1, 8)$ in Figure 4. As we see in Figure 4, P^* and P are very close in this case.

Thus, while Example II.5 illustrated high gains of the proposed code over Shannon codes for P, for the specific case of Zipf distributions the gains may not be large. Characterizing this gain for any given distribution is a direction for future research.

---

[2]We model all our optimization problems in *AMPL* [13] and solve it using *SNOPT* [14] solver.
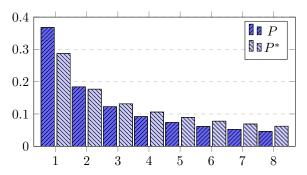
Fig. 4: Pmf for P^* and P for `Zipf`$(1, 8)$.

VI. DISCUSSION ON EXTENSIONS

We have restricted our treatment to deterministic memoryless update schemes. A natural extension to randomized memoryless schemes would entail allowing the encoder to make a randomized decision to skip transmission of a symbol even when the channel is free (we can allocate a special symbol $\emptyset$ to signify no transmission to the receiver). Specifically, denoting by $\theta(x)$ the probability with which the encoder will transmit the symbol x, a modification of the proof of Theorem II.2 yields that the average age for the randomized scheme is given by

$$\frac{\mathbb{E}\left[L(\theta)\right]}{\mathbb{E}\left[\theta(X)\right]} + \frac{\mathbb{E}\left[L(\theta)^2\right]}{2\mathbb{E}\left[L(\theta)\right]} - \frac{1}{2}.$$

Heuristically, it might be baffling what purpose such an omission of transmission can serve. But the following example illustrates the gain in average age that can be obtained by randomly omitting some transmissions.

Example VI.1. Consider $\mathcal{X} = \{1, ..., 64\}$ and the following pmf $P(x) = 1/4$ for $x \in \{1, 2, 3\}$ and $1/244$ otherwise. Since $H(P) = 3.483$, Corollary II.3 yields that the average age of the deterministic memoryless update scheme is bounded below by 4.724. Next, consider a randomized update scheme with $\theta(x) = 1$ for $x \in \{1, 2, 3\}$ and 0 otherwise. For this choice, the effective pmf P_θ is uniformly distributed over the symbols $\{1, 2, 3\} \cup \{\phi\}$. Thus, the optimal length assignment for this case assigns $\ell(x) = 2$ to all the symbols and the average age equals 3.17, which is less than the lower bound of 4.724 for the deterministic scheme.

Next, we point out a use-case for Theorem III.1 in a minimum queuing delay problem introduced in [15]. The setting is similar to the one we described with two differences: First, the arrival process of source symbols is a Poisson process of rate λ; and second, the encoder is not allowed to skip source symbols. Instead, each symbol is encoded and scheduled for transmission in a first-come-first-serve (FCFS) queue. Our goal is to design a source code e that minimizes the average queuing delay $\bar{D}(e)$ encountered by the source sequence. Using the expression for queuing delay for an M/G/1 queue, we can

show that the average delay $\bar{D}(e)$ is given by

$$\bar{D}(e) = \begin{cases} \frac{\mathbb{E}\left[L^2\right]}{2(L_{\text{th}} - \mathbb{E}[L])} + \mathbb{E}\left[L\right], & \mathbb{E}\left[L\right] < L_{\text{th}}, \\ \infty, & \mathbb{E}\left[L\right] \geq L_{\text{th}}, \end{cases}$$

where $L_{\text{th}} := \frac{1}{\lambda}$ is the minimum average codeword length for the queue to be stable. An algorithmic method for finding the length assignments $\ell(x)$, $x \in \mathcal{X}$, that minimizes $\bar{D}(e)$ was presented in [16]. As is clear from the form of average delay formula above, the variational formula of Theorem III.1 and the recipe used to design minimum average age codes in this paper can be used to design minimum average delay codes as well. In simulations we find that our proposed recipe gives performance comparable with [16] when $\mathbb{E}\left[L\right]$ is not very close to L_{th}, namely the moderate load regime, but it requires much less computational effort than the algorithm proposed in [16]. However, we work with rounded-off lengths and do not have a theoretical guarantee of approximation.

Further details for these extensions and the proofs of associated results will be made available in a longer version of this abridged conference submission.

REFERENCES

[1] S. Kaul, R. Yates, and M. Gruteser, "On piggybacking in vehicular networks," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*. IEEE, 2011, pp. 1–5.

[2] ——, "Real-time status: How often should one update?" in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 2731–2735.

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

[4] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems," *IEEE Transactions on Information Theory*, 2017.

[5] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, 2017.

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

[7] A. Mahajan and D. Teneketzis, "Optimal design of sequential real-time communication systems," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5317–5338, 2009.

[8] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1056–1068, 2004.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory. 2nd edition*. John Wiley & Sons Inc., 2006.

[10] L. L. Campbell, "A coding theorem and rényi's entropy," *Information and control*, vol. 8, no. 4, pp. 423–429, 1965.

[11] M. K. Hanawal and R. Sundaresan, "Guessing revisited: A large deviations approach," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 70–78, 2011.

[12] A. Rényi, "On measures of entropy and information," *Proc. Fourth Berkeley Symposium on Mathematics Statistics and Probability, Vol. 1 (Univ. of Calif. Press)*, pp. 547–561, 1961.

[13] R. Fourer, D. M. Gay, and B. Kernighan, *Ampl*. Boyd & Fraser Danvers, MA, 1993, vol. 117.

[14] P. E. Gill, W. Murray, and M. A. Saunders, "Snopt: An sqp algorithm for large-scale constrained optimization," *SIAM review*, vol. 47, no. 1, pp. 99–131, 2005.

[15] P. A. Humblet, "Source coding for communication concentrators," *Ph. D. Dissertation, MIT*, 1978.

[16] L. L. Larmore, "Minimum delay codes," *SIAM Journal on Computing*, vol. 18, no. 1, pp. 82–94, 1989.