Private Status Updating with Erasures: A Case for Retransmission Without Resampling

Ahmed Arafa[1] and Karim Banawan[2]

[1]Electrical and Computer Engineering Department, University of North Carolina at Charlotte, USA
[2]Department of Electrical Engineering, Alexandria University, Egypt

***Abstract*— A status updating system is considered in which a source updates a destination over an erasure channel. The utility of the updates is measured through a function of their *age-of-information* (AoI), which assesses their freshness. Correlated with the status updates is another process that needs to be kept *private* from the destination. Privacy is measured through a *leakage* function that depends on the amount and time of the status updates received: stale updates are more private than fresh ones. Different from most of the current AoI literature, a *post-sampling waiting* time is introduced in order to provide a privacy cover at the expense of AoI. More importantly, it is also shown that, depending on the leakage budget and the channel statistics, it can be useful to retransmit stale status updates following erasure events *without resampling* fresh ones.**

I. Introduction

Providing fresh status updates to destinations is crucial for timely decision-making in various applications, including smart city, e-Health, and digital twins, to name a few. At the same time, with the vast connectivity and correlation between data sources, some information may need to be kept private from curious destinations while providing them with the useful data they need. In this paper, we consider the interplay between utility and privacy of status updates through the lens of *time.*

Data freshness is quantified by the age-of-information (AoI), defined as the time elapsed since the latest useful piece of received data has been generated [1]. In this paper, the utility of status updates is measured through a function acting on their AoI. Such function may represent the estimation mean square error [2], [3], under some assumptions on the underlying process being updated. Privacy, on the other hand, is measured through a function that depends on the relationship between the amount of data received so far and the process to be kept private. We focus on scenarios in which the *privacy leakage is at its peak when status updates are most fresh.* Thereby, a tension arises between data freshness and data privacy.

We study a continuous-time status updating system in which a source-destination pair are communicating through an erasure channel. The freshness of data is controlled by *pre-sampling* waiting times [4], while the privacy is maintained by *post-sampling* waiting times. The post-sampling waiting times are carefully designed to deliver *moderately fresh* updates; these are updates that are fresh enough to provide utility, yet stale enough to provide privacy. A main pillar in our work is that we allow the source to retransmit old samples following erasures without resampling fresh ones. Specifically, we study the following question here:

how many retransmissions are to be allowed before the sample becomes too stale and useless?

We carefully provide an answer to that question that depends on the privacy leakage budget and the channel statistics such that the long-term average utility is maximized.

Related works. A number of works in the literature study the relationship between AoI and privacy. Our previous work [5] considers an information-theoretic private information retrieval problem with AoI guarantees. The work in [6] studies differential privacy metrics that depend on AoI. Reference [7] is closely-related to our work. It considers the privacy-AoI tradeoff in discrete-time systems, and designs post-sample waiting policies for when to release updates in queuing systems in order to control the privacy leakage. Different from [7], we consider a continuous-time system, with erasures, and jointly design waiting times and the number of retransmissions to balance AoI with leakage.

II. System Model and Objective

Consider a stochastic process $\{X_t\}$ that represents a time-varying status to be conveyed to a destination. Such process represents a user's status over time, e.g., home electric usage. Samples from this process are *generated at will,* and are sent through a channel that introduces random delays and erasures. Specifically, the jth sample is generated at time S_j, transmitted for the first time at $T_{j,1}$, and takes $b_{j,1}$ time units to traverse through the channel, denoted the channel *busy time*. After that, the sample is still prune to *erasure* with probability ϵ, whence the sample may be retransmitted at time $T_{j,2}$, incurring $b_{j,2}$ channel busy time, and the process repeats. In general, the jth sample may be (re)transmitted k_j times until successful reception. In case the k_jth attempt fails, the sample is discarded and the process restarts with a *fresh* sample $j+1$. Observe that $k_j = 1$ means that the sample is transmitted only once. We now have the following constraints:

$$T_{j,1} \geq S_j, \quad \forall j, \tag{1}$$

$$T_{j,k+1} \geq T_{j,k} + b_{j,k}, \quad \forall j,\ 1 \leq k \leq k_j, \tag{2}$$

$$S_{j+1} \geq T_{j,k_j} + b_{j,k_j}, \quad \forall j,\ k_j. \tag{3}$$

This work was supported by the U.S. National Science Foundation under Grants CNS 21-14537 and ECCS 21-46099.

Channel busy times, $b_{j,k}$'s, are independent and identically distributed (i.i.d.). Similarly, erasure events are i.i.d., and are independent from $\{X_t\}$ and $b_{j,k}$'s.

A successfully-received sample is denoted an *update*. Let σ_i, τ_i and β_i denote the sampling time of the ith update, its transmission time, and the channel busy time it encounters, respectively. It follows that $\{\sigma_i\} \subseteq \{S_j\}$, $\{\tau_i\} \subseteq \{T_{j,k}\}$ and $\{\beta_i\} \subseteq \{b_{j,k}\}$. The ith update is delivered at time

$$\delta_i = \tau_i + \beta_i. \tag{4}$$

See Fig. 1 for an example time line including all the variables introduced so far. At the destination, the age-of-information (AoI) of the process $\{X_t\}$ at time t is defined as

$$a(t) = t - \max\{\sigma_i : \ \delta_i \le t\}. \tag{5}$$

We measure the utility of the status updates through a general increasing age-penalty functional $g(\cdot)$ that acts upon the AoI process $a(t)$. Specifically, the instantaneous utility of the updates at time t is given by

$$-g\left(a(t)\right). \tag{6}$$

Therefore, updates are more useful when their AoI is small. Observe that the *AoI drops right after delivery times.* We note that measuring utility through AoI is meaningful in estimation and tracking settings, as one can show that the minimum mean square error estimate of Markovian processes is given by an increasing function of the AoI, see, e.g., [2], [3].

Correlated with $\{X_t\}$ is another stochastic process $\{Y_t\}$ that represents a latent variable that needs to be kept *private* from the destination. We consider an *honest-but-curious* destination node that may be interested in getting more information about the user from the updates it conveys. The privacy *leakage* at time t is governed by the amount of information that the received samples, so far, can reveal about Y_t, which we capture using the following non-negative function $\rho(\cdot:\cdot)$:

$$\rho\left(\{X_{\sigma_i}\}_{\delta_i \le t} : Y_t\right), \tag{7}$$

where $\{X_{\sigma_i}\}_{\delta_i \le t}$ denotes all the updates received up to time t. For instance, one can adopt the mutual information [8] to measure the privacy leakage, as done in several works [9]–[16], or other notions such as $\alpha - Leakage$ in [17], [18], and its generalization, $g - Leakage$ in [19]. We have the following assumption about ρ:

$$\rho\left(\{X_{\sigma_i}\}_{\delta_i \le t_1} : Y_{t_1}\right) \ge \rho\left(\{X_{\sigma_i}\}_{\delta_i \le t_2} : Y_{t_2}\right), \\ \forall t_1 < t_2, \text{ s.t. } |\{\delta_i \le t_1\}| = |\{\delta_i \le t_2\}|, \tag{8}$$

where $|\cdot|$ denotes cardinality. Thus, the leakage decreases over time, as long as no new samples have been received. This also implies that *leakage peaks occur right after delivery times.* Several situations satisfy the privacy leakage assumption in (8). For instance, consider the information leakage metric to an estimating adversary in [20],

$$\mathcal{L}(Y_t \to X_t) = \frac{\mathrm{Var}[Y_t]}{\mathbb{E}[(Y_t - \mathbb{E}[Y_t|X_t])^2]}, \tag{9}$$

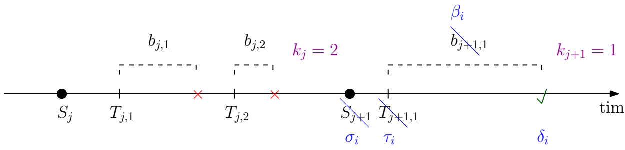

Fig. 1. An example time line evolution; red crosses denote failed transmissions, and the green checkmark denotes a successful transmission.

where the leakage $\mathcal{L}(Y_t \to X_t)$ signifies the estimating accuracy of the adversary (i.e., $\mathcal{L}(Y_t \to X_t) \to \infty$ denotes almost perfect estimation of Y_t given X_t). Considering the estimation setting $X_t = Y_t + W_t$, where W_t is a Wiener process with $W_0 = 0$ and $Y_t \sim \mathcal{N}(0, \sigma_0^2)$ i.i.d. Gaussian process, then $\mathbb{E}[Y_t|X_t] = \frac{\sigma_0^2}{\sigma_0^2+t}X_t,$. The estimation leakage is given by:

$$\rho\left(\{X_{\sigma_i}\}_{\delta_i \le t} : Y_t\right) = \mathcal{L}(Y_t \to X_t) = \frac{1}{1 - \frac{\sigma_0^2}{\sigma_0^2+t}} \tag{10}$$

Hence, $\mathcal{L}(Y_t \to X_t)$ decreases over time. The same arguments hold for the guessing adversary if Y_t is picked from a discrete distribution with $\mathcal{L}(Y_t \to X_t) = \frac{\mathbb{E}[\max_{y \in \mathcal{Y}} P(Y_t|X_t)]}{\max_{y \in \mathcal{Y}} P(Y_t)}$. A different example with an Ornstein-Uhlenbeck (OU) process estimation and a mutual information leakage metric can be found in Section III.

From the above, we see a tension between utility and privacy, as noted in previous works [9]–[12], [14], [16], [18], [21]: *reducing AoI increases the leakage, and vice versa.*

Our goal is to maximize the utility of the updates, while preserving privacy. Specifically, we wish to design the sampling times $\{S_j\}$ and transmission times $\{T_{i,j}\}$ such that the long-term average age-penalty is minimized, subject to an upper bound on the average maximum privacy leakage:

$$\begin{aligned} \min_{\{S_j\},\{T_{j,k}\}} \quad & \limsup_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbb{E}\left[\int_{\delta_{i-1}}^{\delta_i} g\left(t - \sigma_{i-1}\right) dt\right]}{\sum_{i=1}^{n} \mathbb{E}\left[\delta_i - \delta_{i-1}\right]} \\ \text{s.t.} \quad & \mathbb{E}\left[\max_t \rho\left(\{X_{\sigma_i}\}_{\delta_i \le t} : Y_t\right)\right] \le \Delta \\ & (1) - (3), \end{aligned} \tag{11}$$

where $\Delta \ge 0$ denotes the *leakage budget*, and $\mathbb{E}[\cdot]$ denotes the expectation over the random variables involved.

For $\Delta = \infty$, problem (11) reduces to minimizing an AoI functional, and hence it would be optimal to set $T_{j,1} = S_j$ (first transmission times are the same as sampling times) and $k_j = 1$ (only one transmission attempt per sample), $\forall j$. The reason behind these is obvious: there is no need to keep a fresh sample waiting idly *after* being generated as this will only hurt the AoI, and retransmitting an older sample is worse than discarding it and starting fresh. We note that this has been the typical scenario in most AoI literature that do not consider a sampling constraint budget.

Now for $\Delta < \infty$, it may be beneficial to set $T_{j,1} > S_j$. This would relatively increase the AoI of the received update, but at the same time would decrease the privacy leakage. The reason follows from the assumption in (8); the peak leakage occurs at delivery times, and we need to make it no larger than Δ, on average. Following the same rationale, it may also be beneficial to retransmit the same sample in case of a failure since a fresher sample would lead to a relatively lower AoI, and hence a higher privacy leakage, when delivered.

We characterize the solution of (11) in the remainder of this paper. In particular, we discuss when to acquire a new sample, when to transmit it, and how many times it should be retransmitted so as to keep the information about $\{X_t\}$ fresh and that about $\{Y_t\}$ private.

III. Problem Reformulation: Waiting Times and Stationary Policies

In this section, we focus on processes in which the privacy leakage reduces, with a slight abuse of notation, to

$$\begin{aligned}\rho\left(\{X_{\sigma_i}\}_{\delta_i\leq t}:Y_t\right)&\equiv\rho\left(t-\max\left\{\sigma_i:\ \delta_i\leq t\right\}\right)\\&=\rho\left(a(t)\right).\end{aligned}\tag{12}$$

Thus, according to our assumption on ρ, the privacy leakage is a decreasing function of AoI. For instance, consider an OU process $\{X_t\}$ with parameters σ^2 and θ [22], and set $Y_t=X_t+N_t$ with $N_t\sim\mathcal{N}\left(0,\sigma_0^2\right)$ being i.i.d. noise. Further, let the leakage function ρ be given by the mutual information. One can show that [8]

$$\begin{aligned}\rho\left(\{X_{\sigma_i}\}_{\delta_i\leq t}:Y_t\right)&=I\left(\{X_{\sigma_i}\}_{\delta_i\leq t};Y_t\right)\\&=\frac{1}{2}\log\left(\frac{\frac{\sigma^2}{2\theta}+\sigma_0^2}{\frac{\sigma^2}{2\theta}\left(1-e^{-2\theta a(t)}\right)+\sigma_0^2}\right),\end{aligned}\tag{13}$$

which is a decreasing function of $a(t)$ as required.

We now reformulate the optimization problem in (11) in terms of *waiting times,* as opposed to sampling and transmission times. Specifically, we denote by an *epoch* the time elapsed in between two successful updates: the ith epoch extends from δ_{i-1} until δ_i. We now define the first *pre-sampling waiting time* in the ith epoch as

$$W_{i,1}\triangleq\min\left\{S_j:\ S_j\geq\delta_{i-1}\right\}-\delta_{i-1}.\tag{14}$$

That is, $W_{i,1}$ is the waiting time at the beginning of epoch i before acquiring the ***first*** sample in it. Next, we define the first *post-sampling waiting time* following the acquisition of the first sample in the ith epoch as

$$Z_{i,1}^1\triangleq\min\left\{T_{j,1}:\ T_{j,1}\geq\delta_{i-1}\right\}-\left(\delta_{i-1}+W_{i,1}\right).\tag{15}$$

We now (re)denote by $b_{i,1}^k$ the channel busy time of the first sample's kth transmission attempt in the ith epoch. Therefore, we have the post-sampling waiting time sequence given by

$$\begin{aligned}Z_{i,1}^k\triangleq\min&\left\{T_{j,1}:\ T_{j,1}\geq\delta_{i-1}+\sum_{l=1}^{k-1}Z_{i,1}^l+b_{i,1}^l\right\}\\&-\left(\delta_{i-1}+W_{i,1}+\sum_{l=1}^{k-1}Z_{i,1}^l+b_{i,1}^l\right),\quad 1\leq k\leq\kappa_{i,1},\end{aligned}\tag{16}$$

where $\kappa_{i,1}$ now denotes the maximum number of transmission attempts for the first sample in the ith epoch. For simplicity of presentation, let us define

$$M_{i,1}\triangleq\delta_{i-1}+W_{i,1}+\sum_{k=1}^{\kappa_{i,1}}Z_{i,1}^k+b_{i,1}^k\tag{17}$$

as the maximum time allocated to the transmission attempts of the first sample in the ith epoch. In case the first sample is unsuccessful after its last transmission attempt, we define the *second* pre-sampling waiting time as

$$W_{i,2}\triangleq\min\left\{S_j:\ S_j\geq M_{i,1}\right\}-M_{i,1}.\tag{18}$$

This is then followed by a post-sampling waiting time sequence $\{Z_{i,2}^k\}$ given exactly as in (15) and (16) after replacing δ_{i-1}, $W_{i,1}$ and $\kappa_{i,1}$ by $M_{i,1}$, $W_{i,2}$ and $\kappa_{i,2}$, respectively.

In general, the pre- and post-sampling waiting time sequences in the ith epoch will be given as follows:

$$W_{i,r}=\min\left\{S_j:\ S_j\geq M_{i,r-1}\right\}-M_{i,r-1},\tag{19}$$

$$M_{i,r}=M_{i,r-1}+W_{i,r}+\sum_{k=1}^{\kappa_{i,r}}Z_{i,r}^k+b_{i,r}^k,\tag{20}$$

$$\begin{aligned}Z_{i,r}^k=\min&\left\{T_{j,1}:\ T_{j,1}\geq M_{i,r-1}+\sum_{l=1}^{k-1}Z_{i,1}^l+b_{i,1}^l\right\}\\&-\left(M_{i,r-1}+W_{i,r}+\sum_{l=1}^{k-1}Z_{i,r}^l+b_{i,r}^l\right),\quad k\leq\kappa_{i,r},\end{aligned}\tag{21}$$

with $r\geq1$ and $M_{i,0}\triangleq\delta_{i-1}$. The ith epoch ends whenever a transmission attempt is successful, which defines δ_i and the start of the next epoch $i+1$. The length of the ith epoch is

$$\begin{aligned}L_i=\delta_i-\delta_{i-1}=&\sum_{r=1}^{R_i-1}\left(W_{i,r}+\sum_{k=1}^{\kappa_{i,r}}Z_{i,r}^k+b_{i,r}^k\right)\\&+W_{i,R_i}+\sum_{k=1}^{\psi_i}Z_{i,R_i}^k+b_{i,R_i}^k,\end{aligned}\tag{22}$$

where R_i is the number of samples generated in the ith epoch, and ψ_i denotes the number of attempts needed for the R_ith sample to be delivered. Clearly, $\psi_i\leq\kappa_{i,R_i}$. Observe that the AoI at the start of epoch $i+1$ is given by

$$a\left(\delta_i\right)=\sum_{k=1}^{\psi_i}Z_{i,R_i}^k+b_{i,R_i}^k.\tag{23}$$

We focus on *stationary* policies in which the waiting times have the same distribution across epochs. We also fix

$$\kappa_{i,r}=K,\quad\forall i,r.\tag{24}$$

Problem (11) now reduces to one over a single epoch:

$$\begin{aligned}\min_{\{W_{i,r}\geq0\},\ \{Z_{i,r}^k\geq0\},\ K\in\mathbb{Z}_{++}}\quad&\frac{\mathbb{E}\left[\int_0^{L_i}g\left(a\left(\delta_{i-1}\right)+t\right)dt\right]}{\mathbb{E}\left[L_i\right]}\\\text{s.t.}\quad&\mathbb{E}\left[\rho\left(\sum_{k=1}^{\psi_i}Z_{i,R_i}^k+b_{i,R_i}^k\right)\right]\leq\Delta.\end{aligned}\tag{25}$$

We now have the following lemma:

Lemma 1 *In problem (25), it is optimal to perform pre-sampling waiting only at the beginning of an epoch. Likewise, it is optimal to perform post-sampling waiting only once per sample following each sampling time.*

Proof: Observe that the epoch length in (22) only depends on the aggregate sum of the pre-sampling waiting times. One can then define the aggregate pre-sampling waiting time

$$W_i \triangleq \sum_{r=1}^{R_i} W_{i,r}, \tag{26}$$

and optimize that instead, which does not change the value of the optimal solution. Similarly, one can also define aggregate post-sampling waiting times

$$Z_{i,r} \triangleq \sum_{k=1}^{K} Z_{i,r}^k,\ 1 \leq r \leq R_i - 1, \quad Z_{i,R_i} \triangleq \sum_{k=1}^{\psi_i} Z_{i,r}^k, \tag{27}$$

and optimize those instead. ■

Next, we focus on *deterministic waiting policies* in which the pre-sampling waiting time W_i is a deterministic function of the starting AoI of the ith epoch:

$$W_i \equiv w\left(Z_{i-1,R_{i-1}} + \sum_{k=1}^{\psi_{i-1}} b_{i-1,R_{i-1}}^k\right). \tag{28}$$

We note that stationary deterministic waiting policies are known to be optimal under i.i.d. channel settings [4]. By Lemma 1, and under stationary deterministic waiting policies, the ith epoch length is now given by

$$\begin{aligned} L_i =& w\left(Z_{i-1,R_{i-1}} + \sum_{k=1}^{\psi_{i-1}} b_{i-1,R_{i-1}}^k\right) + \sum_{r=1}^{R_i} Z_{i,r} \\ &+ \sum_{r=1}^{R_i-1} \sum_{k=1}^{K} b_{i,r}^k + \sum_{k=1}^{\psi_i} b_{i,R_i}^k, \end{aligned} \tag{29}$$

and the optimization problem finally becomes

$$\begin{aligned} \min_{w(\cdot)\geq 0,\ \{Z_{i,r}\geq 0\},\ K\in\mathbb{Z}_{++}} \quad & \frac{\mathbb{E}\left[\int_0^{L_i} g\left(a\left(\delta_{i-1}\right)+t\right)dt\right]}{\mathbb{E}\left[L_i\right]} \\ \text{s.t.} \quad & \mathbb{E}\left[\rho\left(Z_{i,R_i} + \sum_{k=1}^{\psi_i} b_{i,R_i}^k\right)\right] \leq \Delta. \end{aligned} \tag{30}$$

In the sequel, we present solutions to problem (30) first for the case without errors, followed by that with errors.

IV. The Case Without Errors: $\epsilon = 0$

We analyze problem (30) for error-free transmissions in this section. Our structural insights drawn from the solution of this scenario will serve as a building block for the scenario with channel errors in the following section. Now, for $\epsilon = 0$, we have $K = 1$ (no retransmissions are needed), $R_i = 1,\ \forall i$ and $\psi_i = 1,\ \forall i$. Hence, we drop the r subscript and k superscript in $Z_{i,r}$ and $b_{i,r}^k$, and re-evaluate the epoch length in (29) as

$$L_i = w\left(Z_{i-1} + b_{i-1}\right) + Z_i + b_i. \tag{31}$$

Problem (30) then reduces to

$$\begin{aligned} \min_{w(\cdot)\geq 0,\ \{Z_i\geq 0\}} \quad & \frac{\mathbb{E}\left[\int_0^{L_i} g\left(a\left(\delta_{i-1}\right)+t\right)dt\right]}{\mathbb{E}\left[L_i\right]} \\ \text{s.t.} \quad & \mathbb{E}\left[\rho\left(Z_i + b_i\right)\right] \leq \Delta. \end{aligned} \tag{32}$$

Lemma 2 *The optimal post-sampling waiting policy of problem (32) is $Z_i^* = \zeta,\ \forall i$, for some constant ζ given by*

$$\zeta = \begin{cases} 0, & \textit{if}\ \mathbb{E}\left[\rho\left(b_i\right)\right] < \Delta \\ \xi(\Delta), & \textit{otherwise} \end{cases}, \tag{33}$$

where $\xi(\Delta)$ is the unique solution of $\mathbb{E}\left[\rho\left(\xi(\Delta) + b_i\right)\right] = \Delta$.

Proof: We first argue that Z_i cannot depend on $Z_j,\ j \leq i-1$, since the new sample generated in the ith epoch leaks information at the beginning of epoch $i+1$ with a value that is independent of previous epochs' events. Specifically, Z_i depends solely on b_i (through its distribution). Since b_i's are i.i.d., we can conclude that Z_i's are also i.i.d.

Next, observe that the post-sampling waiting time Z_i can only hurt the AoI. This can readily be shown by a sample path argument; increasing Z_i increases the service time of the ith sample, and only makes it more stale when received. Now let us fix Z_{i-1}. Setting $Z_i = 0$ would then be AoI-optimal, provided that the privacy constraint is met, i.e., if $\mathbb{E}\left[\rho\left(b_i\right)\right] < \Delta$. Otherwise, one should set Z_i to the lowest value allowed by the leakage budget. Since ρ is decreasing, such value is given by $\xi(\Delta)$ in (33).

Finally, since Z_i's are i.i.d., the above argument shows that they should all be fixed at the same value. ■

Lemma 2 shows that there can be situations in which the channel busy time provides a *natural* privacy cover (when $\mathbb{E}\left[\rho\left(b_i\right)\right] < \Delta$), and that post-waiting times should only be used when necessary. The lemma also shows that one can define a *new* channel busy time

$$\tilde{b}_i \triangleq b_i + \zeta, \quad \forall i, \tag{34}$$

with ζ given by (33), which is still i.i.d., and optimize the pre-sampling waiting time over $\{\tilde{b}_i\}$ as done in the AoI minimization literature. Specifically, the results in, e.g., [3], [23], show that the pre-sampling policy is a *threshold policy*

$$w\left(t\right) = \left[G_t^{-1}\left(\gamma\right)\right]^+, \tag{35}$$

where $[\cdot]^+ \triangleq \max\left(\cdot, 0\right)$, and the function

$$G_t\left(x\right) \triangleq \mathbb{E}\left[g\left(t + x + \tilde{b}_i\right)\right] \tag{36}$$

denotes the expected utility by the end of the ith epoch when it starts with an AoI value of t. Further, the value of γ is given by the unique solution of

$$\mathbb{E}\left[\int_0^{\left[G^{-1}_{\tilde{b}_{i-1}}(\gamma)\right]^+ + \tilde{b}_i} g\left(\tilde{b}_{i-1}+t\right)dt\right] - \gamma\mathbb{E}\left[\left[G^{-1}_{\tilde{b}_{i-1}}(\gamma)\right]^+ + \tilde{b}_i\right] = 0, \quad (37)$$

which can be found by, e.g., a bisection search [3].

This completes the solution for the setting without errors.

V. The Case With Errors: $\epsilon > 0$

In this section, we extend the aforementioned solution to the case with channel errors. First, we fix the value of K and evaluate the distributions of the random variables R_i and ψ_i. Observe that a new sample will be generated only if the previous one has K failed transmissions, which occurs with probability ϵ^K. It then follows that

$$R_i \sim \text{geometric}\left(1-\epsilon^K\right), \quad \forall i. \quad (38)$$

As for the number of transmissions, ψ_i, needed for sample R_i (the final sample) to succeed, we note that $\psi_i = k$ in case the kth transmission attempt is successful given that a successful transmission occurs in at most K attempts. Thus,

$$\mathbb{P}\left(\psi_i = k\right) = \frac{\epsilon^{k-1}(1-\epsilon)}{1-\epsilon^K},\ 1 \le k \le K, \quad \forall i, \quad (39)$$

i.e., ψ_i is a truncated $\sim$ geometric$(1-\epsilon)$ random variable.

Next, following similar arguments as in Lemma 2, one can show that the post-sampling waiting times $Z_{i,r}$'s are all fixed in the optimal solution of problem (30), and are given by (33). Hence, the epoch length in (29) is now proportional to

$$R_i\zeta. \quad (40)$$

This allows us to draw the following insight:

Remark 1 *As the privacy leakage budget Δ decreases, ζ increases, and hence the value of R_i must be relatively small so as to not to make the epoch length too large. This can be achieved by increasing K, which controls the distribution of R_i and makes it take smaller values with higher probabilities.*

The above remark is one fundamental observation in this paper: *the number of retransmissions K should be inversely proportional to* Δ. Intuitively, higher levels of privacy are naturally achieved when retransmitting stale samples, and therefore making a case for retransmission without resampling.

To get more insight on how the utility behaves as a function of K, we focus on the scenario in which $g(x) = x$, together with a zero-pre-sampling waiting policy in which $w(t) = 0$. In this case, the starting AoI in the ith epoch is given by

$$\zeta + \sum_{k=1}^{\psi_{i-1}} b^k_{i-1,R_{i-1}}, \quad (41)$$

and the ith epoch length in (29) reduces to

$$L_i = R_i\zeta + \sum_{r=1}^{R_i-1}\sum_{k=1}^{K} b^k_{i,r} + \sum_{k=1}^{\psi_i} b^k_{i,R_i}. \quad (42)$$

Consequently, direct geometrical arguments lead to expressing the utility (long-term average AoI) as

$$\zeta + \mathbb{E}\left[\psi_i\right]\mathbb{E}\left[b_i\right] + \frac{\frac{1}{2}\mathbb{E}\left[L_i^2\right]}{\mathbb{E}\left[R_i\right]\zeta + \mathbb{E}\left[R_i - 1\right]K\mathbb{E}\left[b_i\right] + \mathbb{E}\left[\psi_i\right]\mathbb{E}\left[b_i\right]}. \quad (43)$$

We start with computing the second moment of L_i. Since $\left\{b^k_{i,r}\right\}$, R_i and ψ_i are mutually independent, one can write

$$\begin{aligned}\mathbb{E}\left[L_i^2\right] =& \mathbb{E}\left[\left(R_i\zeta + \sum_{r=1}^{R_i-1}\sum_{k=1}^{K} b^k_{i,r}\right)^2\right] + \mathbb{E}\left[\left(\sum_{k=1}^{\psi_i} b^k_{i,R_i}\right)^2\right] \\ &+ 2\mathbb{E}\left[R_i\zeta + \sum_{r=1}^{R_i-1}\sum_{k=1}^{K} b^k_{i,r}\right]\mathbb{E}\left[\sum_{k=1}^{\psi_i} b^k_{i,R_i}\right]. \quad (44)\end{aligned}$$

One can then show that the second term in (44) is given by

$$\mathbb{E}\left[\left(\sum_{k=1}^{\psi_i} b^k_{i,R_i}\right)^2\right] = \mathbb{E}\left[\psi_i\right]\text{Var}\left[b_i\right] + \mathbb{E}\left[\psi_i^2\right]\left(\mathbb{E}\left[b_i\right]\right)^2, \quad (45)$$

while the first term in (44) can be expressed as

$$\begin{aligned}\mathbb{E}\left[\left(R_i\zeta + \sum_{r=1}^{R_i-1}\sum_{k=1}^{K} b^k_{i,r}\right)^2\right] &= \mathbb{E}\left[R_i^2\right]\zeta^2 \\ +\mathbb{E}\left[\left(\sum_{r=1}^{R_i-1}\sum_{k=1}^{K} b^k_{i,r}\right)^2\right] &+ 2\left(\mathbb{E}\left[R_i^2\right] - \mathbb{E}\left[R_i\right]\right)\zeta K\mathbb{E}\left[b_i\right], \quad (46)\end{aligned}$$

where the last term follows by iterated expectations. Focusing on the second term above, one can expand it as follows:

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{r=1}^{R_i-1}\sum_{k=1}^{K} b^k_{i,r}\right)^2\right] =& \mathbb{E}\left[\sum_{r=1}^{R_i-1}\left(\sum_{k=1}^{K} b^k_{i,r}\right)^2\right] \\ &+ \mathbb{E}\left[\sum_{\substack{r,r'=1 \\ r'\neq r}}^{R_i-1}\left(\sum_{k=1}^{K} b^k_{i,r}\right)\left(\sum_{k'=1}^{K} b^{k'}_{i,r'}\right)\right] \quad (47) \\ =& \mathbb{E}\left[R_i - 1\right]\left(K\mathbb{E}\left[b_i^2\right] + K(K-1)\left(\mathbb{E}\left[b_i\right]\right)^2\right) \\ &+ \mathbb{E}\left[R_i - 1\right]\mathbb{E}\left[R_i - 2\right]K^2\left(\mathbb{E}\left[b_i\right]\right)^2 \quad (48) \\ =& \mathbb{E}\left[R_i - 1\right]K\text{Var}\left[b_i\right] + \left(\mathbb{E}\left[R_i - 1\right]\right)^2 K^2\left(\mathbb{E}\left[b_i\right]\right)^2. \quad (49)\end{aligned}$$

Substituting (49) in (46), and then (46) and (45) in (44), we get an expression for the second moment of L_i in terms of the first and second moments of R_i and ψ_i. These latter moments are directly computable from (38) and (39) as

$$\begin{aligned}\mathbb{E}\left[R_i\right] =& \frac{1}{1-\epsilon^K},\ \mathbb{E}\left[R_i^2\right] = \frac{1+\epsilon^K}{\left(1-\epsilon^K\right)^2} \quad (50) \\ \mathbb{E}\left[\psi_i\right] =& \frac{1-(K+1)\epsilon^K + K\epsilon^{K+1}}{\left(1-\epsilon^K\right)(1-\epsilon)},\end{aligned}$$

$$\mathbb{E}\left[\psi_i^2\right] = \frac{1}{\left(1-\epsilon^K\right)\left(1-\epsilon\right)^2}\left[1+\epsilon-\left(K+1\right)^2\epsilon^K \right. \\ \left. +\left(2K^2+2K-1\right)\epsilon^{K+1}-K^2\epsilon^{K+2}\right]. \quad (51)$$

Finally, observe that the remaining terms in (43) are merely the first moments of R_i and ψ_i computed above. We now have a closed-form expression of the utility in terms of the number of retransmissions K, and the privacy leakage budget Δ (embedded in the value of ζ).

Next, we discuss how to choose the optimal K.

VI. Optimal Number of Retransmissions

We evaluate the optimal number of retransmissions K^* for a system with $b_i \sim \exp(\lambda)$. For the privacy leakage function, we consider the OU process example considered in Section III (ρ is given by (13)). In Fig. 2, we plot the utility versus K for different values of error probability ϵ. In the top figure, we consider a leakage budget of $\Delta = 0.2$, for which one can show by (33) that $\zeta = 0.122$. In the bottom figure, we consider $\Delta = 0.15$, for which $\zeta = 0.24$. In both cases, $\lambda = 10$. Evidently, *the utility in the bottom figure is worse (higher AoI) since the leakage budget is tighter.* Two further observations can be drawn. First, the optimal K^* (circled in red) increases with ϵ for fixed ζ. The intuition behind this is that as ϵ increases, a sample takes a relatively longer time to be successfully delivered. Hence, if one resamples often in this case (i.e., if K is small), the sample will incur even more time due to the extra post-sampling waiting ζ that has to be added. The second observation is that the optimal K^* increases with ζ for fixed ϵ. This is also intuitive since for larger ζ the leakage budget is tighter, and hence one has to retransmit the same sample for longer times to preserve privacy.

Next, we fix $\Delta = 0.2$ (i.e., $\zeta = 0.122$) and vary the busy time statistic, λ. For $\lambda = 1$ (slower channel) we get $K^* = 1$ for all values of ϵ in Fig. 2. While for $\lambda = 20$ (faster channel), we get $K^* = [4, 4, 6, 10]$, i.e., K^* increases relative to the values in Fig. 2. This is mainly because the channel naturally adds a privacy coverage when it is slow, and requires more protection by retranmissions when it is fast.

VII. Conclusion

A freshness-privacy tradeoff has been considered for a source-destination pair communicating through an erasure channel. It has been shown that carefully designing *post-sampling waiting times,* together with the *number of retransmissions of unsuccessful samples* can provide useful status updates while maintaining desired levels of privacy.

References

[1] R. D. Yates, Y. Sun, R. D. Brown, S. K. Kaul, E. Modiano, and S. Ulukus. Age of information: An introduction and survey. *IEEE J. Sel. Areas Commun.*, 39(5):1183–1210, May 2021.

[2] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu. Sampling of the Wiener process for remote estimation over a channel with random delay. *IEEE Trans. Inf. Theory*, 66(2):1118–1135, February 2020.

[3] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor. Sample, quantize, and encode: Timely estimation over noisy channels. *IEEE Trans. Commun.*, 69(10):6485–6499, October 2021.

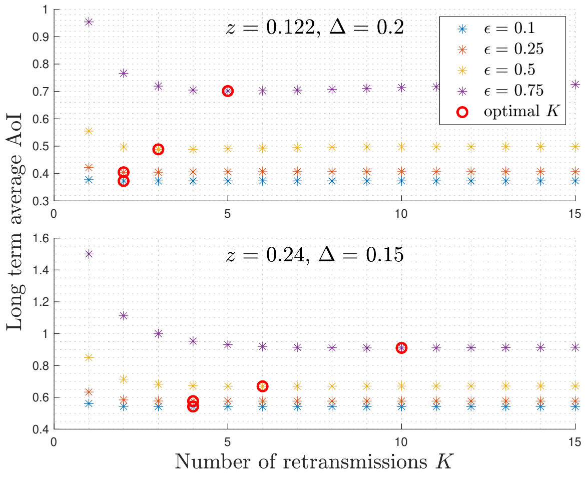

Fig. 2. Utility vs. retransmissions; channel busy time $\sim \exp(10)$.

[4] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff. Update or wait: How to keep your data fresh. *IEEE Trans. Inf. Theory*, 63(11):7492–7508, November 2017.

[5] K. Banawan, A. Arafa, and S. Ulukus. Timely private information retrieval. In *Proc. IEEE ISIT*, July 2021.

[6] M. Zhang, E. Wei, R. Berry, and J. Huang. Age-dependent differential privacy. In *Proc. ACM Sigmetrics/IFIP Performance*, June 2022.

[7] N. Sathyavageesran, R. D. Yates, A. D. Sarwate, and N. Mandayam. Privacy leakage in discrete-time updating systems. In *Proc. IEEE ISIT*, June 2022.

[8] T. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2006.

[9] L. Sankar, S. Kar, R. Tandon, and H. V. Poor. Competitive privacy in the smart grid: An information-theoretic approach. In *Proc. IEEE SmartGridComm*, October 2011.

[10] F. P. Calmon and N. Fawaz. Privacy against statistical inference. In *Proc. Allerton*, October 2012.

[11] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Medard. From the information bottleneck to the privacy funnel. In *Proc. IEEE ITW*, November 2014.

[12] S. Asoodeh, F. Alajaji, and T. Linder. On maximal correlation, mutual information and data privacy. In *Proc. IEEE CWIT*, July 2015.

[13] W. Wang, L. Ying, and J. Zhang. On the relation between identifiability, differential privacy, and mutual-information privacy. *IEEE Trans. Inf. Theory*, 62(9):5018–5029, September 2016.

[14] J. Liao, L. Sankar, V. F. Tan, and F. P. Calmon. Hypothesis testing under mutual information privacy constraints in the high privacy regime. *IEEE Trans. Inf. Forensics Security*, 13(4):1058–1071, April 2018.

[15] S. Li, A. Khisti, and A. Mahajan. Information-theoretic privacy for smart metering systems with a rechargeable battery. *IEEE Trans. Inf. Theory*, 64(5):3679–3695, May 2018.

[16] S. Sreekumar and D. Gunduz. Optimal privacy-utility trade-off under a rate constraint. In *Proc. IEEE ISIT*, July 2019.

[17] Jiachun J. Liao, L. Sankar, O. Kosut, and F. P. Calmon. Maximal α-leakage and its properties. In *Proc. IEEE CNS*, June 2020.

[18] A. Kamatsuka, T. Yoshida, and T. Matsushima. Privacy-utility trade-off with the Stratonovich's value of information. In *Proc. IEEE ITW*, October 2021.

[19] G. Kurri, L. Sankar, and O. Kosut. An operational approach to information leakage via generalized gain functions. arXiv:2209.13862.

[20] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder. Estimation efficiency under privacy constraints. *IEEE Trans. Inf. Theory*, 65(3):1512–1534, March 2019.

[21] S. R. Rajagopalan, L. Sankar, S. Mohajer, and H. V. Poor. Smart meter privacy: A utility-privacy framework. In *Proc. IEEE SmartGridComm*, October 2011.

[22] G. E. Uhlenback and L. S. Ornstein. On the theory of the Brownian motion. *Phys. Rev.*, 36:823–841, September 1930.

[23] Y. Sun and B. Cyr. Sampling for data freshness optimization: Non-linear age functions. *J. Commun. Netw.*, 21(3):204–219, June 2019.