Trading off SNR and the Number of Observations to Improve the Value of Information in IoT Networks

Zijing Wang, Mihai-Alin Badiu, and Justin P. Coon
Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, United Kingdom
e-mail: {zijing.wang, mihai.badiu, justin.coon}@eng.ox.ac.uk

***Abstract*—The freshness and usefulness of information play an important role in offering ubiquitous connectivity for time-critical control applications. A concept named value of information (VoI) is proposed based on the field of information theory to quantify the usefulness of data for sensor-assisted Internet of Things (IoT) networks in the presence of transmission noise. In this work, we focus on general Gaussian random process models and study the rate of change of the VoI when generating more data samples and increasing the signal-to-noise ratio (SNR). We further look at Gauss-Markov random process models, and investigate the impact of the number of observations and the SNR on the VoI performance. It is interesting to find that using more data samples is effective to improve the VoI only in the low SNR regime, while it yields zero rate of change of the VoI in the high SNR regime. Moreover, the VoI can be improved by increasing the SNR in both high and low SNR regimes regardless of how many samples are used. We also find a trade-off between the SNR and the number of observations, and scale back SNR to achieve the same VoI improvement by adding one extra observation. The results illustrated in this work can be used in the design of practical real-time IoT networks.**

***Index Terms*—Value of information, data freshness, Internet of Things**

I. Introduction

Driven by the rapid advance in intelligent sensor-based Internet of Things (IoT), fresh and useful information has become increasing important in order to make real-time decisions and guarantee efficient network operation. Sensing devices are deployed to continuously collect the ambient data in IoT, and the data is required to be transmitted to a remote monitor as timely as possible to make sure that the monitor always holds the latest status information of a target physical process. A metric named age of information (AoI) is employed to measure the freshness of information from the receiver's side [1]. The AoI is defined as the time duration since the generating time of the latest received data sample. It increases linearly with unit slope and decreases suddenly when a new packet is observed.

The research on AoI has received increasingly interest over the recent years [2]. The AoI has been linked to queueing theory [3], signal processing theory (estimation theory) [4] and information theory [5], and it has worked as an effective tool in the design of freshness-aware network architectures [6]. AoI-optimal scheduling and wireless resource allocation problems have been extensively investigated in a variety of specific 5G/6G scenarios [7]–[9]. Furthermore, AoI variants have also been introduced to meet different demands in different applications. Since the performance of the AoI degrades linearly over time, non-linear AoI functions have been explored to quantify how dissatisfying the outdated information is [10]–[12].

Despite these contributions, the effects of the signal-to-noise ratio (SNR) and the correlation between data samples on the data freshness have not been clearly treated. In practice, samples can be negatively affected if they are observed through a poor channel, thus the usefulness of data deteriorates. However, the quality of the transmission channel was ignored in many existing works. Moreover, existing AoI and its variants characterise the data freshness by observing the most recent sample. Using a single observation is reasonable in non-noisy Markov models, but in the presence of noise, all the past observations at the receiver are correlated and they all contain valuable information about the underlying data source. Therefore, both the SNR and past data samples are important factors affecting the data freshness. Regarding this background, we are motivated to employ the AoI in an information-theoretic context, and study how the usefulness of data changes with the SNR and the number of past observations.

Recently, we proposed the concept of the *value of information* (VoI) based on the mutual information and obtained its closed-form expression in general Gaussian process models [13], [14]. In this work, we explore the impact of the SNR and the number of observations on the VoI with the aim of improving its performance. Specifically, we study the rate of change of the VoI with respect to the number of observations and the SNR in Gaussian models. We further apply these results to Gauss-Markov models in which the correlation function is exponential. We find that, in the high SNR regime, increasing the SNR is a possible method to improve the VoI, but using more data samples is not. In the low SNR regime, both methods can help to increase the VoI, and there exists a trade-off: either increasing SNR for a fixed number of observations or using an extra data sample for a fixed SNR can achieve the same target value. The results presented in this work can be applied in various Gauss-Markov models, such as the autoregressive model and the Ornstein–Uhlenbeck model, to improve the usefulness of information in practical IoT-assisted control applications.

II. System Model

A. Value of Information Definition

We consider a sensor-based IoT system in which sensing devices are deployed to monitor the surrounding environment, traffic and so on. The physical phenomenon under observation is modelled as a random process $\{X_t\}$. The information generated by the sensor at the transmitter needs to be communicated to a remote monitor promptly, and the processed data will be further used to control a plant in such real-time monitoring systems. We denote X_{t_i} as the data packet which is sampled at time t_i. It represents the status of the underlying random process at t_i and i is the index of this sample. From the receiver's perspective, $\{Y_t\}$ is the observed random process. We denote $Y_{t'_i}$ as the i-th corresponding observation which is received at time t'_i. Due to the transmission and queuing delay, we assume that $t'_i > t_i$. Due to the noise and interference through the transmission channel, we assume that $\{Y_t\} \neq \{X_t\}$ which means that the underlying random process is not directly observable at the receiver.

Given a time instant t, the traditional age of information concept is defined as $t - t_n$ in which $t > t'_n$. X_{t_n} represents the latest received data sample and the AoI only measures the time duration since the most recent received packet is generated. In our work, we define the value of information in an information-theoretic context. Specifically, the VoI is defined as the mutual information between the state of the underlying random process at a given time t and the latest m observations, i.e.,

$$v(t) = \mathrm{I}(X_t; Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}), \quad t > t'_n. \tag{1}$$

Here, n is the total number of data samples observed by the monitor and we extract the latest m ($1 \leq m \leq n$) observations to measure the usefulness of the data. This mutual information-based VoI notion measures the reduction in uncertainty for the current unknown status of the underlying latent process given a collection of observed measurements. If the entropy of X_t is identical regardless of the time (i.e., the random process $\{X_t\}$ is stationary), the VoI can be functionally linked to the conditional entropy of X_t given $\{Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}\}$. In this case, the VoI can also be interpreted as how much information the observation $\{Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}\}$ tells us about a latent status X_t.

B. Noisy Gaussian Process Model

We denote vector $\boldsymbol{X}_m$ and $\boldsymbol{Y}_m$ as the sets of the latest m samples and observations, respectively, i.e., $\boldsymbol{X}_m = [X_{t_n}, \cdots, X_{t_{n-m+1}}]^{\mathrm{T}}$ and $\boldsymbol{Y}_m = [Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}]^{\mathrm{T}}$. We assume that the stochastic process $\{X_t\}$ is Gaussian and stationary. Let the variance of this Gaussian process be σ_x^2 and the mean be 0. For any $1 \leq i \leq n$, X_{t_i} is normally distributed with $X_{t_i} \sim \mathcal{N}(0, \sigma_x^2)$. $\boldsymbol{X}_m$ is multivariate normally distributed with $\boldsymbol{X}_m \sim \mathcal{N}(\boldsymbol{0}, \sigma_x^2 \mathbf{K}_{\boldsymbol{X}_m})$ where the autocorrelation matrix $\mathbf{K}_{\boldsymbol{X}_m}$ is given by

$$\mathbf{K}_{\boldsymbol{X}_m} = \begin{bmatrix} 1 & r_{t_n - t_{n-1}} & \cdots & r_{t_n - t_{n-m+1}} \\ r_{t_n - t_{n-1}} & 1 & \cdots & r_{t_{n-1} - t_{n-m+1}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{t_n - t_{n-m+1}} & r_{t_{n-1} - t_{n-m+1}} & \cdots & 1 \end{bmatrix}. \tag{2}$$

In this matrix, r is the Pearson correlation coefficient which is used to represent the dependency between the data samples. Given a pair of random variables (X_{t_i}, X_{t_j}), it can be presented by $r_{|t_i - t_j|}$, which only relates to the time difference, for stationary random processes. The correlation coefficient ranges from -1 to 1. If $r = 0$, there is no linear correlation between two variables. If $r = 1$ or -1, they can be regarded as positively or negatively highly-correlated variables.

Data samples are communicated through the following additive white Gaussian noise channel

$$\boldsymbol{Y}_m = \boldsymbol{X}_m + \boldsymbol{N}_m \tag{3}$$

where $\boldsymbol{N}_m = [N_{t'_n}, \cdots, N_{t'_{n-m+1}}]^{\mathrm{T}}$. We denote $\{N_{t'_i}\}$ as noise samples which are independently, identically and normally distributed with $N_{t'_i} \sim \mathcal{N}(0, \sigma_n^2)$ for any $1 \leq i \leq n$. In this model, we denote γ as the signal-to-noise ratio (SNR) which is given by

$$\gamma = \frac{\sigma_x^2}{\sigma_n^2}. \tag{4}$$

III. VoI in Gaussian Models

In the model we described above, the value of information defined in (1) for general Gaussian and stationary process models is given in the following lemma [14].

Lemma 1. *For a fixed time instant t, the VoI for the noisy Gaussian model is given as*

$$v(m, \gamma) = -\frac{1}{2} \log \left[1 - \boldsymbol{u}_m^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma} \mathbf{I}_m \right)^{-1} \boldsymbol{u}_m \right], \tag{5}$$

where the vector $\boldsymbol{u}_m$ is given by

$$\boldsymbol{u}_m = \left[r_{t-t_n}, r_{t-t_{n-1}}, \cdots, r_{t-t_{n-m+1}} \right]^{\mathrm{T}}. \tag{6}$$

Here, $\mathbf{K}_{\boldsymbol{X}_m}$ is the autocorrelation matrix which is given in (2) and $\mathbf{I}_m$ is the m-dimensional identity matrix. Vector $\boldsymbol{u}_m$ measures the dependency between the state at time t and past states of the underlying random process. The VoI is presented by a function of m and γ. m is the number of observations are used and the SNR γ represents the quality of the communication channel. Compared with the traditional AoI concept, it is easy to see that this VoI concept measures both the correlation property of the underlying data source[1] and whether the channel condition is poor or good.

Lemma 1 shows that the VoI is largely affected by the number of past observations m and the SNR γ. For simplicity, we denote η as the absolute value of the exponential VoI difference, i.e.,

$$\eta = \left| e^{-2v(m,\gamma)} - e^{-2v(m',\gamma')} \right|. \tag{7}$$

η is a non-negative function of m, m', γ and γ'. It can be used to represent the growth of the amount of valuable information

[1]The correlation can be affected by either the inherent property of the underlying data source (which depends on the autocorrelation function r in r_δ) or the time duration between the generation time of two samples (which depends on the time difference δ in r_δ).

by using more past observations or increasing the SNR. We consider the following two cases to further explore the VoI: (1) For a fixed SNR γ, the change of the VoI when increasing m by one, which is denoted as η_m; (2) for a fixed number of observations m, the change of the VoI when increasing a certain value of the SNR, which is denoted as η_γ. Then, we can state the following two propositions.

Proposition 1. *For a fixed time instant t and a fixed SNR γ, the exponential VoI difference over the number of observations ($m' = m+1$) can be given by*

$$\eta_m = \rho\left[\boldsymbol{u}_m^{\mathrm{T}}\left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma}\mathbf{I}_m\right)^{-1}\boldsymbol{w}_m - r_{t-t_{n-m}}\right]^2. \quad (8)$$

Vector $\boldsymbol{w}_m$ is given by

$$\boldsymbol{w}_m = \left[r_{t_n-t_{n-m}}, r_{t_{n-1}-t_{n-m}}, \cdots, r_{t_{n-m+1}-t_{n-m}}\right]^{\mathrm{T}}. \quad (9)$$

ρ is given by

$$\rho = \left[1 + \frac{1}{\gamma} - \boldsymbol{w}_m^{\mathrm{T}}\left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma}\mathbf{I}_m\right)^{-1}\boldsymbol{w}_m\right]^{-1}. \quad (10)$$

Proof: See Appendix A. ∎

This proposition shows the improvement of the VoI when increasing the number of observations by 1 for a fixed SNR, i.e., $\eta_m = e^{-2v(m,\gamma)} - e^{-2vv(m+1,\gamma)}$. $r_{t-t_{n-m}}$ captures the correlation between the current status X_t and the $(m+1)$-th measurement $Y_{t'_{n-m}}$.

Proposition 2. *For a fixed time instant t and a fixed m, the exponential VoI difference over the SNR can be given by*

$$\eta_\gamma = \boldsymbol{u}_m^{\mathrm{T}}\left[\left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma'}\mathbf{I}_m\right)^{-1} - \left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma}\mathbf{I}_m\right)^{-1}\right]\boldsymbol{u}_m. \quad (11)$$

Proposition 2 shows the improvement of the VoI when increasing the SNR from γ to γ' for a fixed m, i.e., $\eta_\gamma = e^{-2v(m,\gamma)} - e^{-2v(m,\gamma')}$ where $\gamma' > \gamma$.

IV. VoI in Gauss-Markov Models

In this section, we apply the results in Propositions 1 and 2 to a special case in which the underlying source data is exponentially correlated. If the underlying stochastic process $\{X_t\}$ is Gaussian and its autocorrelation function is an exponential function, then $\{X_t\}$ is not only a Gaussian but also a first-order Markov process. This case can be used to model many practical Gauss-Markov random processes, such as the autoregressive process and the Ornstein–Uhlenbeck process. We focus on this general Gauss-Markov model to further study how the VoI varies with the number of observations and the SNR.

We write the correlation coefficient as $r_\delta = \beta^\delta$ where β is a constant, ranging from 0 to 1. We assume that samples are generated at arbitrary but fixed times $\{t_i\}$, and denote τ_i as the sampling interval of two packets, i.e.,

$$\tau_i = t_{n+1-i} - t_{n-i}, \quad 1 \le i \le m. \quad (12)$$

In the absence of the noise (when $\sigma_n^2 = 0$), the underlying random process is not latent from the receiver's perspective (i.e., $X_{t_i} = Y_{t'_i}$), thus the VoI defined in (1) can be written as

$$v(t) = \mathrm{I}(X_t; X_{t_n}) = -\frac{1}{2}\log\left[1 - \beta^{2(t-t_n)}\right], \quad t > t'_n. \quad (13)$$

In this case, the VoI does not relate to the number of observations or the quality of the transmission channel. It only relates to the correlation function and the time elapsed until a new observation is made.

In the presence of the noise, it is easy to see that

$$\mathrm{I}(X_t; Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}) \le \mathrm{I}(X_t; X_{t_n}). \quad (14)$$

The equality holds when the SNR $\gamma \to \infty$ or $\sigma_n^2 = 0$. In this case, the following corollaries can be derived from Propositions 1 and 2 under different SNR conditions.

A. High SNR regime

Corollary 1. *In the high SNR regime, for a fixed time instant t and a fixed SNR γ, the exponential VoI difference over the number of observations satisfies*

$$\eta_m = O(\gamma^{-m}). \quad (15)$$

Moreover, $\eta_m = 0$ holds in the absence of the noise.

Proof: See Appendix B. ∎

Corollary 1 shows how the VoI varies when increasing m by 1 for a fixed and high SNR. For the exponentially correlated data source, the underlying random process is Markov and the observed process will also turn to be Markov in the absence of noise. Therefore, in Corollary 1, the difference between the VoI with $m+1$ and m observations is presented by a big O term in which γ^{-m} approaches 0 exponentially quickly as m grows. This means that increasing the length of the observation window does not help to improve the VoI in this case. Regardless of the correlation condition, the high SNR regime yields nearly zero rate of change of the VoI when increasing the number of observations.

Corollary 2. *In the high SNR regime, for a fixed time instant t and a fixed m, the exponential VoI difference over the SNR satisfies*

$$\eta_\gamma = \left(\frac{1}{\gamma} - \frac{1}{\gamma'}\right)\beta^{2(t-t_n)} + O\left(\frac{1}{\gamma^2}\right) + O\left(\frac{1}{\gamma'^2}\right). \quad (16)$$

Proof: In this case, we assume that $\frac{1}{\gamma} \to 0$ and $\frac{1}{\gamma'} \to 0$. This result is obtained directly by $(\mathbf{K}_{\boldsymbol{X}_m} + \epsilon\mathbf{I}_m)^{-1} = \mathbf{K}_{\boldsymbol{X}_m}^{-1} - \epsilon(\mathbf{K}_{\boldsymbol{X}_m}^{-1})^2 + O(\epsilon^2)$ when $\epsilon \to 0$ in Proposition 2. ∎

Corollary 2 shows the change of the VoI when increasing the SNR for a fixed m. The increase of the VoI with different SNR does not relate to m, thus the SNR is the dominant factor affecting the VoI in the high SNR regime. When $m = 1$, the difference of the VoI in Corollary 1 can be presented as the same order in Corollary 2. This means that, only for very small m, we may scale back SNR as we increase m by one to achieve the same VoI.

B. Low SNR regime

Corollary 3. *In the low SNR regime, for a fixed time instant t and a fixed SNR γ, the exponential VoI difference over the number of observations satisfies*

$$\eta_m = \beta^{2(t-t_n)}\beta^{2\sum_{i=1}^{m}\tau_i}\left[\gamma - \left(1+2m\right)\gamma^2\right] + O(\gamma^3). \quad (17)$$

Proof: See Appendix C. ■

In the low SNR regime, Corollary 3 shows that increasing the number of observations is a possible way to improve the VoI. However, η_m decreases with increasing m. This means that infinite observations cannot increase the VoI infinitely. Moreover, the low correlation condition ($\beta \to 0$) also yields a zero rate of change of the VoI when increasing m which is the same as the high SNR regime. A highly correlated data source ($\beta \to 1$) yields a positive VoI difference.

Corollary 4. *In the low SNR regime, for a fixed time instant t and a fixed m, the exponential VoI difference over the SNR satisfies*

$$\eta_\gamma = (\gamma' - \gamma)\left(1 + \sum_{j=1}^{m-1}\beta^{2\sum_{i=1}^{j}\tau_i}\right)\beta^{2(t-t_n)} + O(\gamma^2) + O(\gamma'^2). \quad (18)$$

Proof: This results are obtained by series expansion at $\gamma = 0$ and $\gamma' = 0$. ■

Corollary 4 shows that increasing the SNR is another possible way to improve the VoI in the low SNR regime. There is a trade-off existing in this case, i.e., either increasing SNR or increasing m by one can achieve the same target value. Compared with Corollaries 3 and 4, we have the following remark.

Remark 1. *Increasing the SNR by a certain amount $\Delta\gamma$ has the same effect on the VoI as increasing m by 1 in the low SNR regime. Specifically, for fixed m and γ, we have $v(m+1,\gamma) = v(m, \gamma + \Delta\gamma)$ when*

$$\Delta\gamma = \frac{\gamma}{\sum_{j=1}^{m}\beta^{-2\sum_{i=j}^{m}\tau_i}}. \quad (19)$$

V. Numerical Results

Numerical results obtained by Monte Carlo simulations are provided in this section. In the simulation, data samples are generated randomly at rate λ according to a Poisson process. The transmission delay of each sample is also generated randomly according to an exponential distribution with rate μ. We set $\lambda = 0.5$ and $\mu = 1$, and evaluate the VoI at a fixed time instant $t = 100$.

Fig. 1 shows the normalised VoI with the different number of observations m and different SNR γ, and illustrates the

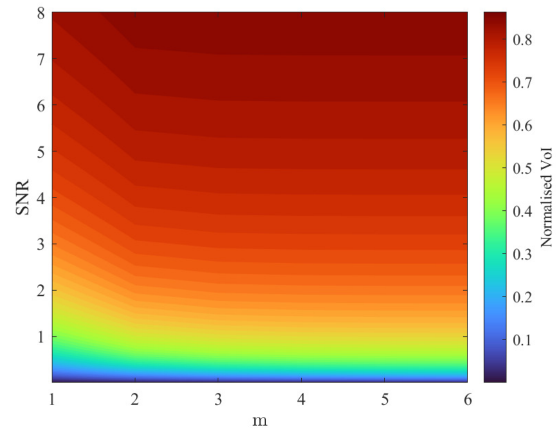

Fig. 1. The normalised value of information with different number of observations $m \in \{1,2,3,4,5,6\}$ and different SNR $\gamma \in (0,8)$. The correlation parameter $\beta = 0.7$.

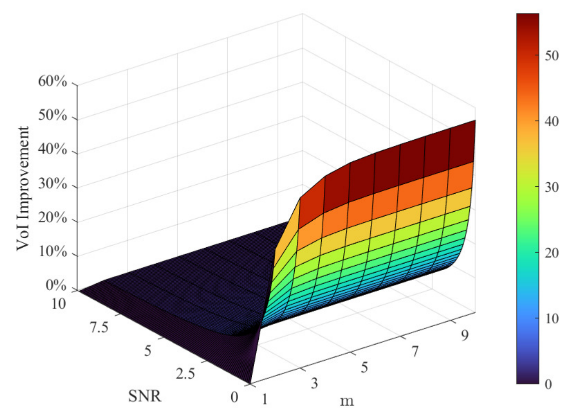

Fig. 2. The improvement of the value of information by increasing the number of observations $m \in \{1,2,3,\cdots,8,9,10\}$ under different SNR conditions $\gamma \in (0,10)$. The correlation parameter $\beta = 0.7$.

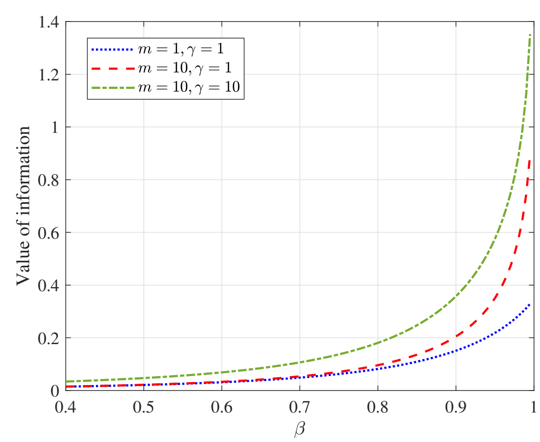

Fig. 3. The value of information versus the correlation parameter $\beta \in (0.4, 1)$.

trade-off between m and γ. The normalised VoI is the ratio of the VoI and its non-noisy Markov counterpart which is given in (13). It is shown that the VoI increases and then converges when increasing the SNR or using more data samples. Fixing the SNR, the VoI increases with m when the SNR is small, while the change is not obvious when the SNR is large. This result verifies the statements given in Corollaries 1 and 3, i.e., only in the low SNR regime, the VoI can be increased by increasing m. Fixing m, the VoI increases with the SNR regardless of the number of observations. This means that increasing the SNR is helpful to increase the VoI and verifies the results given in Corollaries 2 and 4.

Fig. 2 further shows that the performance of the VoI when m rises from 1 to 10 for different SNR. The vertical axis represents the percentage increase in the VoI which is calculated by $\frac{v(m,\gamma)-v(1,\gamma)}{v(1,\gamma)} \times 100\%$. When the SNR is low, the VoI can be improved up to 60% by using more past observations. When the SNR is high, large m is not helpful to improve the VoI and the SNR is the dominant term affecting the VoI.

Fig. 3 illustrates that how the VoI varies with the autocorrelation parameter β. Large β represents the highly-correlated data source and small β means weak correlation. Regardless of the SNR and the number of observations, the VoI always increases with the correlation parameter. As β rises, the past data samples will be more dependent and contain more valuable information about the underlying random process, which further yields higher VoI.

VI. Conclusions

In this work, we focused on general Gaussian process models and investigated how the SNR and the number of observations affect the value of information and how to improve the performance of the VoI. We further looked at an exponentially decaying autocorrelation function and studied the trade-off between the two in different SNR conditions. It is interesting to find that, in the high SNR regime, increasing the SNR can largely increase the VoI regardless of how many observations are captured. However, using more past data samples yields nearly zero rate of change of the VoI in this case. In the low SNR regime, we showed that increasing the SNR by a certain value can achieve the equivalent VoI performance as increasing the number of observations by 1. The analytical results obtained in this work can be further used to optimise the value of information for time-critical Gauss-Markov models.

Appendix A
Proof of Proposition 1

Let the vector

$$\boldsymbol{w}_m = \left[r_{t_n - t_{n-m}}, r_{t_{n-1} - t_{n-m}}, \cdots, r_{t_{n-m+1} - t_{n-m}} \right]^{\mathrm{T}}.$$

The autocorrelation matrix of $\boldsymbol{X}_{m+1}$ can be written as

$$\mathbf{K}_{\boldsymbol{X}_{m+1}} = \begin{bmatrix} \mathbf{K}_{\boldsymbol{X}_m} & \boldsymbol{w}_m \\ \boldsymbol{w}_m^{\mathrm{T}} & 1 \end{bmatrix}. \tag{20}$$

For simplicity, we denote

$$\boldsymbol{\Sigma}_{m+1} = \mathbf{K}_{\boldsymbol{X}_{m+1}} + \frac{1}{\gamma}\mathbf{I}_{m+1} = \begin{bmatrix} \boldsymbol{\Sigma}_m & \boldsymbol{w}_m \\ \boldsymbol{w}_m^{\mathrm{T}} & 1+\frac{1}{\gamma} \end{bmatrix}. \tag{21}$$

The correlation between the current status X_t and the past $m+1$ observations can be captured by the following vector

$$\boldsymbol{u}_{m+1} = [\boldsymbol{u}_m^{\mathrm{T}}, r_{t-t_{n-m}}]^{\mathrm{T}}. \tag{22}$$

Based on Lemma 1, the VoI with $m+1$ observations is given by

$$v(m+1,\gamma) = -\frac{1}{2}\log\left[1 - \boldsymbol{u}_{m+1}^{\mathrm{T}} \boldsymbol{\Sigma}_{m+1}^{-1} \boldsymbol{u}_{m+1}\right]. \tag{23}$$

Based on the block matrix inversion, we have

$$\boldsymbol{\Sigma}_{m+1}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_m^{-1} + \rho \boldsymbol{\Sigma}_m^{-1}\boldsymbol{w}_m \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} & -\rho \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m \\ -\rho \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} & \rho \end{bmatrix} \tag{24}$$

where

$$\rho = \left(1 + \frac{1}{\gamma} - \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m\right)^{-1} = \frac{\gamma}{1+\gamma-\gamma \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m}. \tag{25}$$

Therefore, the VoI difference can be presented by

$$\begin{aligned} &\boldsymbol{u}_{m+1}^{\mathrm{T}} \boldsymbol{\Sigma}_{m+1}^{-1} \boldsymbol{u}_{m+1} \\ &\quad = \boldsymbol{u}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{u}_m + \rho\left(\boldsymbol{u}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m - r_{t-t_{n-m}}\right)^2. \end{aligned} \tag{26}$$

Appendix B
Proof of Corollary 1

For exponentially correlated random processes, we have

$$\boldsymbol{u}_m = \beta^{t-t_n}\left[1, \beta^{\tau_1}, \cdots, \beta^{\sum_{i=1}^{m-1}\tau_i}\right]^{\mathrm{T}}, \tag{27}$$

$$\boldsymbol{w}_m = \beta^{\tau_m}\left[\beta^{\sum_{i=1}^{m-1}\tau_i}, \beta^{\sum_{i=2}^{m-1}\tau_i}, \cdots, 1\right]^{\mathrm{T}}, \tag{28}$$

$$r_{t-t_{n-m}} = \beta^{t-t_n}\beta^{\sum_{i=1}^{m}\tau_i}, \tag{29}$$

$$\mathbf{K}_{\boldsymbol{X}_m} = \begin{bmatrix} 1 & \beta^{\tau_1} & \cdots & \beta^{\sum_{i=1}^{m-1}\tau_i} \\ \beta^{\tau_1} & 1 & \cdots & \beta^{\sum_{i=2}^{m-1}\tau_i} \\ \vdots & \vdots & \ddots & \vdots \\ \beta^{\sum_{i=1}^{m-1}\tau_i} & \beta^{\sum_{i=2}^{m-1}\tau_i} & \cdots & 1 \end{bmatrix}, \tag{30}$$

and

$$\mathbf{K}_{\boldsymbol{X}_m}^{-1} = \begin{bmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & b_2 & \ddots & \ddots & \\ & & \ddots & a_{m-1} & b_{m-1} \\ & & & b_{m-1} & a_m \end{bmatrix}. \tag{31}$$

Here,

$$a_i = \begin{cases} \frac{1}{1-\beta^{2\tau_1}} & i=1 \\ \frac{1}{1-\beta^{2\tau_{m-1}}} & i=m \\ \frac{1}{1-\beta^{2\tau_{i-1}}} + \frac{1}{1-\beta^{2\tau_i}} - 1 & \text{others} \end{cases} \tag{32}$$

and

$$b_i = -\frac{\beta^{\tau_i}}{1-\beta^{2\tau_i}}, \quad 1 \le i \le m-1. \tag{33}$$

In the high SNR regime, we assume that $\frac{1}{\gamma} \to 0$. Series expansion is used at $\frac{1}{\gamma} = 0$ to study how the VoI varies with the length of the time window m. For simplicity, we denote $\frac{1}{\gamma}$ as z and let $\eta(z) = \frac{g^2(z)}{h(z)}$ where

$$\begin{aligned} g(z) &= \boldsymbol{u}_m^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}_m} + z\mathbf{I}_m \right)^{-1} \boldsymbol{w}_m - r_{t-t_{n-m}} \\ h(z) &= 1 + z - \boldsymbol{w}_m^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}_m} + z\mathbf{I}_m \right)^{-1} \boldsymbol{w}_m. \end{aligned} \tag{34}$$

It is easy to see that $g(0) = 0$ and $g'(0) = 0$, thus we have $\eta(0) = 0$ and $\eta'(0) = 0$.

For any positive integer $k > 1$, k-th order derivatives of $g(z)$ and $h(z)$ are given as

$$\frac{\partial^k g}{\partial z^k} = (-1)^k \boldsymbol{u}_m^{\mathrm{T}} \left[\left(\mathbf{K}_{\boldsymbol{X}_m} + z\mathbf{I}_m \right)^{-1} \right]^{k+1} \boldsymbol{w}_m \prod_{i=1}^{k} i, \tag{35}$$

$$\frac{\partial^k h}{\partial z^k} = (-1)^{k+1} \boldsymbol{w}_m^{\mathrm{T}} \left[\left(\mathbf{K}_{\boldsymbol{X}_m} + z\mathbf{I}_m \right)^{-1} \right]^{k+1} \boldsymbol{w}_m \prod_{i=1}^{k} i. \tag{36}$$

Since

$$\boldsymbol{u}_m^{\mathrm{T}} \mathbf{K}_{\boldsymbol{X}_m}^{-1} = \beta^{t-t_n} [1, 0, \cdots, 0] \tag{37}$$

$$\mathbf{K}_{\boldsymbol{X}_m}^{-1} \boldsymbol{w}_m = \beta^{\tau_m} [0, \cdots, 0, 1]^{\mathrm{T}}, \tag{38}$$

we have $\frac{\partial^k g}{\partial z^k} = 0$ holds for any $1 \le k \le m-1$ when $z = 0$. Therefore, we can state that $\frac{\partial^k \eta}{\partial z^k} = 0$ at $z = 0$ for any $1 \le k \le m-1$.

If $\frac{1}{\gamma} = 0$ we have $\boldsymbol{u}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m = r_{t-t_{n-m}}$ in Proposition 1, thus the VoI with $m+1$ observations equals its counterpart with m observations when there is no noise.

Appendix C
Proof of Corollary 3

Similar to the analysis given in the high SNR regime, we let $\eta(\gamma) = \gamma \frac{g^2(\gamma)}{h(\gamma)}$ where

$$\begin{aligned} g(\gamma) &= \gamma \boldsymbol{u}_m^{\mathrm{T}} \left(\gamma \mathbf{K}_{\boldsymbol{X}_m} + \mathbf{I}_m \right)^{-1} \boldsymbol{w}_m - r_{t-t_{n-m}} \\ h(\gamma) &= 1 + \gamma - \gamma^2 \boldsymbol{w}_m^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}_m} + \mathbf{I}_m \right)^{-1} \boldsymbol{w}_m. \end{aligned} \tag{39}$$

It is easy to see that $g(0) = -r_{t-t_{n-m}}$ and $h(0) = 1$. The derivatives of $g(\gamma)$ and $h(\gamma)$ are given as

$$g'(0) = \boldsymbol{u}_m^{\mathrm{T}} \boldsymbol{w}_m, \quad g''(0) = -2\boldsymbol{u}_m^{\mathrm{T}} \mathbf{K}_{\boldsymbol{X}_m}^{-1} \boldsymbol{w}_m \tag{40}$$

$$h'(0) = 1, \quad h''(0) = 2\boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{w}_m. \tag{41}$$

Therefore, derivatives of $\eta(\gamma)$ are given as

$$\begin{aligned} \eta'(0) &= r_{t-t_{n-m}}^2, \\ \eta''(0) &= -2r_{t-t_{n-m}}^2 - 4m r_{t-t_{n-m}} \beta^{t-t_n} \beta^{\sum_{i=1}^{m} \tau_i}. \end{aligned} \tag{42}$$

Acknowledgements

This work was supported by EPSRC grant number EP/T02612X/1. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising. The authors also gratefully acknowledge the support of the Clarendon Fund Scholarships of the University of Oxford.

References

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

[2] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, “Age of information: An introduction and survey,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, 2021.

[3] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, “A general formula for the stationary distribution of the age of information and its application to single-server queues,” *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8305–8324, 2019.

[4] 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,” in *Proc. IEEE INFOCOM Workshops*, 2018, pp. 367–372.

[5] Y. Sun and B. Cyr, “Information aging through queues: A mutual information perspective,” in *Proc. IEEE SPAWC*, 2018, pp. 1–5.

[6] A. Kosta, N. Pappas, and V. Angelakis, *Age of Information: A New Concept, Metric, and Tool*. Now Foundations and Trends, 2017.

[7] M. A. Abd-Elmagid and H. S. Dhillon, “Average peak age-of-information minimization in UAV-assisted IoT networks,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 2003–2008, 2019.

[8] Z. Wang, X. Qin, B. Liu, and P. Zhang, “Joint data sampling and link scheduling for age minimization in multihop cyber-physical systems,” *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 765–768, 2019.

[9] Q. He, D. Yuan, and A. Ephremides, “Optimal link scheduling for age minimization in wireless systems,” *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5381–5394, 2018.

[10] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, “The cost of delay in status updates and their value: Non-linear ageing,” *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4905–4918, 2020.

[11] X. Zheng, S. Zhou, and Z. Niu, “Urgency of information for context-aware timely status updates in remote control systems,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7237–7250, 2020.

[12] Y. Sun and B. Cyr, “Sampling for data freshness optimization: Non-linear age functions,” *J. Commun. Netw.*, vol. 21, no. 3, pp. 204–219, 2019.

[13] Z. Wang, M.-A. Badiu, and J. P. Coon, “A framework for characterising the value of information in hidden Markov models,” *IEEE Trans. Inf. Theory*, 2022.

[14] ——, “On the value of information and mean squared error for noisy Gaussian models,” *submitted to IEEE Commun. Lett.*, 2022.