Information-Centric Scheduling for Wireless Sensor Networks by Adaptive-Rate Compression

Chi-Jen Lo* and Hung-Yun Hsieh*†
*Graduate Institute of Communication Engineering
†Department of Electrical Engineering
National Taiwan University, Taipei, Taiwan 106
Email: {R08942126, hungyun}@ntu.edu.tw

***Abstract*—In this paper, we consider the problem of minimizing the queue length for data gathering in wireless sensor networks (WSNs) while considering the quality of information received at the sink. We extend the nominated Slepian-Wolf-Cover bound for distributed source coding (DSC) into blockwise streaming scheme, where distributed sources can be compressed without loss exploiting the spatio-temporal correlations between data samples from individual sensors. Most importantly, unlike the classic Slepian-Wolf-Cover bound, which requires the statistical information of the information sources *a priori*, our proposed time-average inequality can dynamically adapts to the information in every scheduling period. We contrive and solve an optimization problem to show the capability of this bound in real applications.**

I. Introduction

Information entropy, defined in *information theory*, measures the information content and can be considered as the minimum amount of bits to compress data (and remove redundancy) without resulting in information loss [1]. *Information-centric communications* which prioritize information over devices for allocating communication resources, is in contrast to conventional "device-centric" communications which treat information from different devices as independent bit streams, aiming to optimize the *quality of service (QoS)* of individual devices such as transmission rate, delay, and outage under constrained resource [2]–[4]. Nevertheless, optimizing QoS such as the sum rate or throughput fairness could result in reception of redundant information, causing the waste of communication resources. In scenarios like *wireless sensor networks (WSNs)*, since the sensors are deployed to collectively gather data, data from different devices often exhibits correlation [5]. To better serve these kind of applications, it is the optimization of the quality of information received that should be focused on.

In this vein, different information-centric mechanisms that aim to optimize the fidelity of received information set by applications have been investigated [6]–[10]. Although these endeavors have substantiated several benefits of information-centric communications, most have so far explored only spatial or temporal correlation among information gathered by different sensor nodes [11], [12]. In this paper, however, we explore both spatial and temporal correlations of the data [13]–[15]. As a classic element in information theory, *distributed source coding (DSC)* regards the compression of multiple correlated information sources that do not communicate with each other. Modeling correlation between multiple sources at the decoder side, Slepian-Wolf-Cover bound provides the theoretical guarantee to losslessly shift the computational complexity from encoders to the joint decoder, thus makes appropriate frameworks for applications like sensor networks and multimedia compression [16]. While several works have applied Slepian-Wolf-Cover bound on resource allocation problems [17], [18], we extend the classic Slepian-Wolf-Cover source coding theorem and propose another information-theoretic bound that allows compression rate to vary in different scheduling periods. This extension allows the scheduler to decide and allocate the communication resources period by period, without knowing the statistical information of the sources in advance. Such adaptive property coincides with the scheduling mechanism of wireless sensor networks aggregating stochastic information of interests from the environment [19], [20]. Therefore, we formulate an optimization problem on minimize queue length scheduling for wireless sensor networks indicating the practical use of our proposed time-average inequality. We then transform this bound into virtual queues, and represent the rate region by matrices assuming multivariate Gaussian distribution [1], [21].

II. Slepian-Wolf-Cover in Blockwise Streaming

A. Source Model

We denote $\mathcal{S} = \{1, 2, 3, \dots, |\mathcal{S}|\}$ as the set of sensor nodes where each $i \in \mathcal{S}$ generates data samples as triggered by the event of interests, and the data samples generated by sensor i at different timestamps follow an arbitrary ergodic process, denoted by X_i. T represents the length of a time period for aggregating the sampled data. At the end of each period, all samples are aggregated and compressed into one data packet, waiting for transmission in the sensor queue. $X_i[k]$ is the random vector with length $|X_i[k]| = n_i[k]$, consisting of data samples gathered from time $t = kT$ to $t = (k+1)T$, represented by stochastic process $X_i[k] = \{X_i(t) \,|\, \forall\, kT \leq t < (k+1)T\}$. Hence, the amount of data stored by sensor i in its queue at time $t = (k+1)T$ (i.e. at the end of period k) is $C_i[k] = H(X_i[k])$. The queue size of sensor i at time $t = kT$ (i.e. the beginning of period k) is $Q_i[k]$. Queue is initially empty with $Q_i[0] = 0$ and the value of $Q_i[1]$ is equal to the amount of data $C_i[0]$, aggregated from $t = 0$ to $t = T$.

B. Slepian-Wolf-Cover Bound in Blockwise Streaming Scheme

Given a general quantity of discrete memoryless independent channels with multiple supply nodes and a single sink, the *Slepian-Wolf-Cover theorem* extends the original Slepian-Wolf coding to multiple sources [22]. The Slepian-Wolf-Cover bound can be written as:

$$\begin{aligned}\sum_{i\in W}\gamma_i &\geq H(X_W|X_{W^c}) = H(X_{W\cup W^c}) - H(X_{W^c}) \\ &= H(X) - H(X_{W^c}), \forall\, W\subseteq \mathcal{S}, W^c = \mathcal{S}-W,\end{aligned} \tag{1}$$

where γ_i is the coding rate of i and $H(X_W)$ is the entropy of $\cup_{i\in W}X_i$. Such bound indicates the theoretical rate region for the lossless compression of multiple sources, yet there is a major bottleneck that plagues its practical use. That is, the statistical information among all information sources has to be revealed and known *a priori*. In this research, we leverage this theorem for compressing multiple correlated information sources to come up with a new time-average bound, which allows the system to make decision of the coding rate period by period, leading to the reduction of average queue length.

In other words, we regard to a blockwise streaming scheme, where time period T represents a block, and the compression rate of each sensor node adapts at every time period k. This is nonetheless different from the original Slepian-Wolf-Cover theorem as in (1), so we propose Theorem 1 as an extension:

Theorem 1 (The Slepian-Wolf-Cover Bound for Blockwise Streaming Scheme)**.** *Denote $\gamma_i[k]$ as the compression rate of sensor i in kth scheduling period, then the extended Slepian-Wolf-Cover bound for the blockwise streaming scheme is:*

$$\lim_{L\to\infty}\frac{1}{L}\sum_{k=1}^{L}\sum_{i\in W}\gamma_i[k] \geq \lim_{L\to\infty}\frac{1}{L}\sum_{k=1}^{L}\Big[H(X_W[k]|X_{W^c}[k])\Big], \tag{2}$$

for all $W\subseteq\mathcal{S}$, $W^c=\mathcal{S}-W$. Concatenating the data sample vector $X_i[k]$ from all sensors together in the order of their indexes, we obtain $X[k]$, which can be seen as a sequence keeping all the sampled information in scheduling period k. Since ergodicity, this is equivalent to the following inequality:

$$\lim_{L\to\infty}\frac{1}{L}\sum_{k=1}^{L}\Big[\mathbb{E}\big[H(X_W[k]|X_{W^c}[k])\big] - \mathbb{E}\big[\sum_{i\in W}\gamma_i[k]\big]\Big] \leq 0, \tag{3}$$

which guarantees the condition for lossless data compression.

Proof. Starting from a single source model, the *Asymptotic Equipartition Property (AEP)* indicates that while $\mathcal{A}_\epsilon^{(n)}$ is a set of typical sequences, then $|\mathcal{A}_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$ for any $\epsilon > 0$. Suppose we have $f: X^n \to (1,2,...,2^{n\gamma})$ and $f(x^n)$ is a random index from $\{1,2,...,2^{n\gamma}\}$, applying the random binning technique, the error probability of decoding this sequence would be:

$$\begin{aligned}P_e^{(n)} &= Pr[\exists\, x^n\neq y^n, x^n\in\mathcal{A}_\epsilon^{(n)}, y^n\in\mathcal{A}_\epsilon^{(n)}, f(x^n)=f(y^n)] \\ &\leq \sum_{y^n\in\mathcal{A}_\epsilon^{(n)}} Pr[y^n] \sum_{x^n\in\mathcal{A}_\epsilon^{(n)}, x^n\neq y^n} Pr[f(x^n)=f(y^n)] \\ &= \sum_{y^n\in\mathcal{A}_\epsilon^{(n)}} Pr[y^n] \sum_{x^n\in\mathcal{A}_\epsilon^{(n)}} 2^{-n\gamma} \leq 2^{-n\gamma}2^{n(H(X)+\epsilon)}.\end{aligned} \tag{4}$$

Hence, for source with $n\to\infty$, the error probability $P_e^{(n)}$ falls to 0 while the compression rate $\gamma \geq H(X)+\epsilon$, $\forall\epsilon>0$.

After the example, we can now generalize multiple sources lossless coding into blockwise streaming scheme. To illustrate, we show the distributed lossless compression of two ergodic sources X_i and X_j. By *Shannon-McMillan-Breiman theorem*,

$$\begin{aligned}&-\frac{1}{L}\log p(X_i[1],...,X_i[L],X_j[1],...,X_j[L]) \to H(X_i,X_j), \\ &-\frac{1}{L}\log p(X_i[1],...,X_i[L]) \to H(X_i), \\ &-\frac{1}{L}\log p(X_j[1],...,X_j[L]) \to H(X_j)\end{aligned} \tag{5}$$

with probability one, where

$$\begin{aligned}H(X_i,X_j) &= \lim_{L\to\infty}\frac{1}{L}H(X_i[1],...,X_i[L],X_j[1],...,X_j[L]), \\ H(X_i) &= \lim_{L\to\infty}\frac{1}{L}H(X_i[1],...,X_i[L]), \\ H(X_j) &= \lim_{L\to\infty}\frac{1}{L}H(X_j[1],...,X_j[L]).\end{aligned} \tag{6}$$

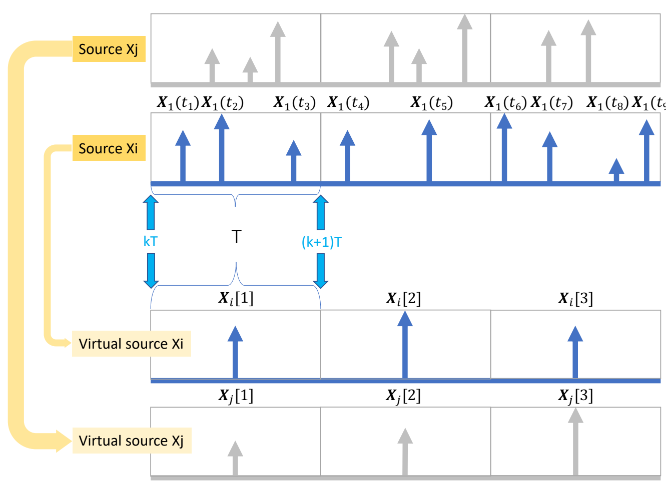

Fig. 1: Illustration of how we associate multiple arbitrary ergodic sources into correspondence with those L-sequences and the predefined entropy rates. In each period, all samples within which are aggregated and represented by a new signal.

Consequently, as illustrated in Fig. 1, for arbitrary ergodic sources X_i, X_j and $\epsilon>0$, there exists integer L and a set

$\mathcal{T}_\epsilon^{(L)}$ of typical L-sequences $\boldsymbol{x_i} = (x_i[1], ..., x_i[L])$, $\boldsymbol{x_j} = (x_j[1], ..., x_j[L])$ such that

$$\begin{aligned} p(\boldsymbol{x_i}, \boldsymbol{x_j}) &= p(\overbrace{x_i[1], ..., x_i[L]}^{L}, \overbrace{x_j[1], ..., x_j[L]}^{L}) \\ &= p(\overbrace{x_i(t_1), ..., x_i(t_{n_i})}^{n_i=\sum_{k=1}^{L} n_i[k]}, \overbrace{x_j(t_1), ..., x_j(t_{n_j})}^{n_j=\sum_{k=1}^{L} n_j[k]}) \end{aligned} \tag{7}$$

and

$$Pr[\mathcal{T}_\epsilon^{(L)}] \equiv Pr[(\boldsymbol{x_i}, \boldsymbol{x_j}) \in \mathcal{T}_\epsilon^{(L)}] = \sum_{\mathcal{T}_\epsilon^{(L)}} p(\boldsymbol{x_i}, \boldsymbol{x_j}) \geq 1 - \epsilon. \tag{8}$$

Therefore,

$$\begin{aligned} \left| -\frac{1}{L} \log p(\boldsymbol{x_i}, \boldsymbol{x_j}) - H(X_i, X_j) \right| &\leq \frac{\epsilon}{2}, \\ \left| -\frac{1}{L} \log p(\boldsymbol{x_i}) - H(X_i) \right| &\leq \frac{\epsilon}{2}, \\ \left| -\frac{1}{L} \log p(\boldsymbol{x_j}) - H(X_j) \right| &\leq \frac{\epsilon}{2}, \end{aligned} \tag{9}$$

where the first inequality in (9) implies

$$2^{-L(H(X_i,X_j)+\epsilon/2)} \leq p(\boldsymbol{x_i}, \boldsymbol{x_j}) \leq 2^{-L(H(X_i,X_j)-\epsilon/2)}. \tag{10}$$

Additionally, $H(X_i|X_j) \leq \lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} H(X_i[k]|X_j[k])$ holds and it can be derived from

$$\begin{aligned} &\lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} H(X_i[k]|X_j[k]) \\ &= \lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} \Big[H(X_i[k], X_j[k]) - H(X_j[k]) \Big] \\ &= \lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} H(X_i[k], X_j[k]) - \lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} H(X_j[k]) \\ &\geq H(X_i, X_j) - \lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} H(X_j[k]) \\ &\geq H(X_i, X_j) - H(X_j) = H(X_i|X_j), \end{aligned} \tag{11}$$

given that $\lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} H(X_i[k], X_j[k]) \geq H(X_i, X_j)$ and $H(X_j) \leq \lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} H(X_j[k])$. The second inequality can be proven by applying Hadamard-Fischer's inequality on (26) whereas the first inequality can be proven along these lines:

$$\begin{aligned} &\lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} H(X_i[k], X_j[k]) \\ &= \lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} \Big[-\sum_{\boldsymbol{x_i}, \boldsymbol{x_j}} p(x_i[k], x_j[k]) \log p(x_i[k], x_j[k]) \Big] \\ &= \lim_{L\to\infty} \sum_{\boldsymbol{x_i}, \boldsymbol{x_j}} \log \Bigg[\prod_{k=1}^{L} \left(\frac{1}{p(x_i[k], x_j[k])} \right)^{p(x_i[k], x_j[k])/L} \Bigg] \\ &\geq \lim_{L\to\infty} \sum_{\boldsymbol{x_i}, \boldsymbol{x_j}} \log \Bigg[\prod_{k=1}^{L} \left(\frac{1}{p(x_i[k], x_j[k])} \right)^{\prod_{k=1}^{L} p(x_i[k], x_j[k])} \Bigg] \\ &= \lim_{L\to\infty} \sum_{\boldsymbol{x_i}, \boldsymbol{x_j}} \log \Bigg[\left(\frac{1}{p(\boldsymbol{x_i}, \boldsymbol{x_j})} \right)^{p(\boldsymbol{x_i}, \boldsymbol{x_j})} \Bigg] = H(X_i, X_j). \end{aligned} \tag{12}$$

Now, let the set of $\boldsymbol{x_i}$ that are *jointly typical* with $\boldsymbol{x_j}$ be defined by $\mathcal{J}_{\boldsymbol{x_j}} = \{\boldsymbol{x_i} \,|\, (\boldsymbol{x_i}, \boldsymbol{x_j}) \in \mathcal{T}_\epsilon^{(L)}\}$. Since for each $\boldsymbol{x_j}$

$$\begin{aligned} 1 &= \sum_{\boldsymbol{x_i}} p(\boldsymbol{x_i}|\boldsymbol{x_j}) = \sum_{\boldsymbol{x_i}} \frac{p(\boldsymbol{x_i}, \boldsymbol{x_j})}{p(\boldsymbol{x_j})} \geq \sum_{\boldsymbol{x_i} \in \mathcal{J}_{\boldsymbol{x_j}}} \frac{p(\boldsymbol{x_i}, \boldsymbol{x_j})}{p(\boldsymbol{x_j})} \\ &\geq \sum_{\boldsymbol{x_i} \in \mathcal{J}_{\boldsymbol{x_j}}} \frac{2^{-L(H(X_i,X_j)+\epsilon/2)}}{2^{-L(H(X_j)-\epsilon/2)}} = |\mathcal{J}_{\boldsymbol{x_j}}| 2^{-L(H(X_i|X_j)+\epsilon)}, \end{aligned} \tag{13}$$

using the concept of random binning as in (4) and (11) yields

$$\begin{aligned} P_e^{(L)} &\leq \lim_{L\to\infty} \sum_{(\boldsymbol{x_i}, \boldsymbol{x_j}) \in \mathcal{T}_\epsilon^{(L)}} p(\boldsymbol{x_i}, \boldsymbol{x_j}) |\mathcal{J}_{\boldsymbol{x_j}}| \Big[\prod_{k=1}^{L} 2^{-\gamma_i[k]} \Big] \\ &\leq \lim_{L\to\infty} 2^{L(H(X_i|X_j)+\epsilon)} 2^{-\sum_{k=1}^{L} \gamma_i[k]} \\ &\leq \lim_{L\to\infty} 2^{-L\left(\frac{1}{L} \sum_{k=1}^{L} \gamma_i[k] - \frac{1}{L} \sum_{k=1}^{L} H(X_i[k]|X_j[k]) - \epsilon \right)} = 0 \end{aligned} \tag{14}$$

under condition $\lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} \big[\gamma_i[k] - H(X_i[k]|X_j[k]) \big] > 0$. Similarly, this proof can be extended to multiple sources. □

To interpret, Theorem 1 loosens the original Slepian-Wolf-Cover bound (1), that in some periods, the compression rates can be outside of the achievable rate region if ergodically the time-average compression rate remains in the time-average achievable rate region. For notation simplicity, we applied the standard notation concerning conditional entropies (Gallager, 1968) [23], and let $\sigma_W[k] = H(X_W[k]|X_{W^c}[k])$. We can also let $\sum_{i\in W} \gamma_i[k] = \gamma_W[k]$, then (3) is furthermore shorten as:

$$\lim_{L\to\infty} \frac{1}{L} \sum_{k=1}^{L} \mathbb{E}\big[\sigma_W[k] - \gamma_W[k] \big] \leq 0, \forall\, W \subseteq \mathcal{S}, W^c = \mathcal{S} - W. \tag{15}$$

Moreover, to avoid the separate decoding scheme, we require

$$0 < \gamma_i[k] \leq H(X_i[k]),\ \forall k \in \mathbb{N}, \forall\, i \in \mathcal{S}. \tag{16}$$

III. An Application and A Special Case of Gaussian

In this section, we contrive a time-average resource allocation problem, which actually originates the idea on bound (2) as an example. [1] This example also leads to an outer bound assuming multivariate Gaussian. In scheduling period k, each sensor queue $i \in \mathcal{S}$ distributively compress its selected data for uploading by $\gamma_i[k]$ and then proactively drop $d_i[k]$ amount of data remaining, constrained by the minimum information fidelity requirement. Considering the simplest TDMA scheme, the scheduler allocates proper scheduling time ratio $\tau_i[k]$ for those sensors. $R_i[k]$ is the transmission rate in physical layer.

A. *Objective, Problem Transformation and Evaluations*

At each scheduling period $k \in \mathbb{N}$, the scheduler determines the transmission time ratio $\tau_i[k]$, compression rate $\gamma_i[k]$ and proactively dropped data $d_i[k]$ for each sensor $i \in \mathcal{S}$ to minimize the average queue length of all sensors, written as:

$$\underset{\{\boldsymbol{\tau}[k],\boldsymbol{\gamma}[k],\mathbf{d}[k]\}}{\text{Minimize}} \lim_{L\to\infty} \frac{1}{L}\sum_{k=1}^{L}\sum_{i\in\mathcal{S}} Q_i[k]. \tag{17}$$

Having the objective function (17) and the extended Slepian-Wolf-Cover bound (2), we can formulate an optimization problem for resource allocation over L periods. Yet there is no way for us to directly minimize (17) since this problem belongs to a stochastic optimization problem. To deal with these kind of problems, we apply the theoretical tool of Lyapunov optimization for transforming the problem [9], [24].

By minimizing the drift $\Delta(\boldsymbol{\Theta}[k])$ in each scheduling period, which is identical to maximizing (18) through choosing the decision vector $\{\boldsymbol{\tau}[k],\boldsymbol{\gamma}[k],\boldsymbol{d}[k]\}$ at each time period k, we can minimize the time-average queue length while ensuring queue stability. Inequality (2) is changed into virtual queues (24) whereas the mean rate stability of the virtual queues validates the satisfaction of the condition for lossless distributed source coding (i.e. the extended Slepian-Wolf-Cover bound) and the desired information fidelity. $Q_i[k]C_i[k]$ and $V_W[k]\sigma_W[k]$ are removed due to their independence of the decision variables. After the transformation, we obtain a series of Markov decision (MDP) problems as in Table I. These problems belong to signomial geometric programming (SGP), which are mostly NP-hard. Along with the method in [25], a branch-and-bound algorithm is provided for attaining the global optimal.

Although the model is eligible for arbitrary ergodic source, to construct an example, we assume $X_i[k]$ follows a multivariate Gaussian distribution. This assumption is quite acceptable since Gaussian drives the largest differential entropy among all kinds of distributions, and can be seen as the worst case. We denote the covariance matrix as $\Sigma_{X_i}[k]$, and its determinant as $\det(\Sigma_{X_i}[k])$. Assuming a uniform quantizer of step size Δ is applied to all data samples, and therefore the entropy (i.e. the minimum amount of bits needed to encode/compress the source) of the random vector can be approximated as in [1],

$$H(X_i[k]) \approx \frac{1}{2}\log_2\left[\frac{(2\pi e)^{|X_i[k]|}}{\Delta^{|X_i[k]|}}\det(\Sigma_{X_i}[k])\right]. \tag{26}$$

[1]Please resort to [9] for detailed introduction of the communication model.

TABLE I: Formulation for Information-Centric Scheduling

Objective:

$$\sum_{i\in S} Q_i[k](\tau_i[k]TR_i[k]/\gamma_i[k] + d_i[k]) + \sum_{W\subseteq S} V_W[k]\gamma_W[k] + \sum_{i\in S} Z_i[k][(1-\xi)\tau_i[k]TR_i[k] - \xi d_i[k]] \tag{18}$$

Subject to:

Non-negative time ratio constraint:

$$\tau_i[k] \geq 0, \forall i \in \mathcal{S}, \tag{19}$$

Transmission time summation constraint:

$$\sum_{i=1}^{M} \tau_i[k] \leq 1, \tag{20}$$

Non-negative dropped data constraint:

$$d_i[k] \geq 0, \forall i \in \mathcal{S}, \tag{21}$$

Non-empty queue constraint:

$$Q_i[k] \geq \tau_i[k]TR_i[k]/\gamma_i[k] + d_i[k], \forall i \in \mathcal{S}, \tag{22}$$

Upper bound on compression rate:

$$0 < \gamma_i[k] \leq H(X_i[k]), \forall i \in \mathcal{S}, \tag{16}$$

Queue update:

$$Q_i[k+1] = \max\{Q_i[k] - \tau_i[k]TR_i[k]/\gamma_i[k] - d_i[k], 0\} + C_i[k], \forall i \in \mathcal{S}, \tag{23}$$

Virtual queue update for lossless distributed source coding:

$$V_W[k+1] = \max\{V_W[k] - \sum_{i\in W}\gamma_i[k] + \sigma_W[k], 0\}, \forall W \subseteq \mathcal{S}, \tag{24}$$

Virtual queue update ensuring information fidelity ξ:

$$Z_i[k+1] = \max\{Z_i[k] + \xi d_i[k] - (1-\xi)\tau_i[k]TR_i[k], 0\}, \forall i \in \mathcal{S}. \tag{25}$$

We also define the spatio-temporal covariance matrix necessary for quantifying term $\sigma_W[k]$ for multivariate Gaussian,

$$\begin{aligned}\sigma_W[k] &= H(X_W[k]|X_{W^c}[k]) = H(X[k]) - H(X_{W^c}[k]) \\ &\approx \frac{1}{2}\log_2\left[\frac{(\frac{2\pi e}{\Delta})^{|X[k]|}\det(\Sigma_X[k])}{(\frac{2\pi e}{\Delta})^{|X_{W^c}[k]|}\det(\Sigma_{X_{W^c}}[k])}\right].\end{aligned} \tag{27}$$

Fig. 2 exhibits the covariance matrix for $X[k]$, denoted as $\Sigma_X[k]$, where $|X[k]| = n[k]$ is the total number of samples in period k. Since $X[k]$ is concatenated by $X_i[k]$, $n[k] = \sum_{i\in\mathcal{S}} n_i[k]$. Simulation results in Fig. 3 shows our proposed information-centric scheduling policy can significantly improve the performance by reducing the average queue length.

B. *Exponential Reduction of the Virtual Queues*

Strongly due to the nature of Gaussian entropy region [21], $2^{|\mathcal{S}|}$ virtual queues V_W are needed for storing the information of (24), causing impractical space complexity. Thus, we propose Theorem 2 to exponentially reduce the quantity of virtual queues and derive an outer bound of (2) in terms of matrices:

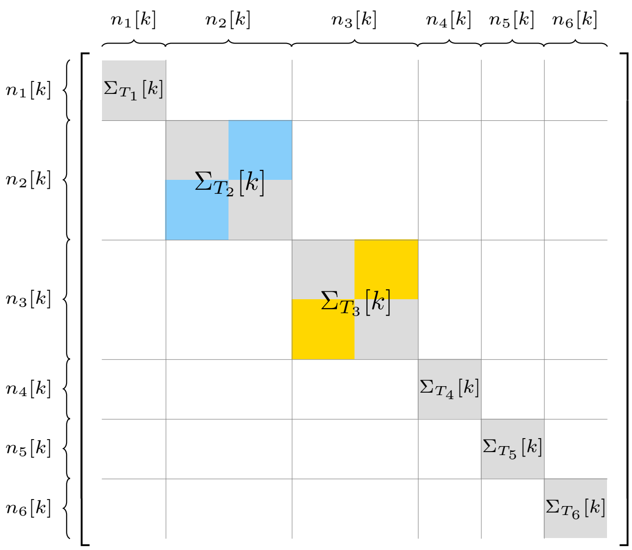

Fig. 2: Structure of the spatio-temporal covariance matrix. The diagonal blocks store the temporal covariances between samples while the off-diagonal blocks store the spatial covariances.

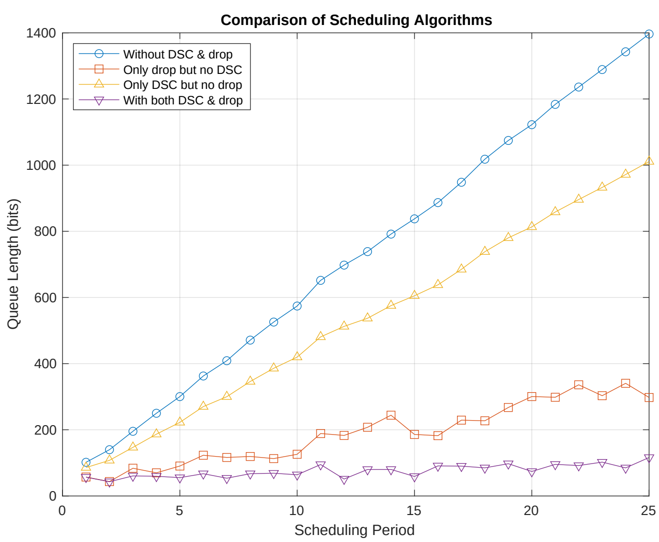

Fig. 3: Comparison of scheduling algorithms ($\xi = 0.95$)

Theorem 2. *The exponentially many (w.r.t. $|\mathcal{S}|$) virtual queues proposed in* (24) *can be rearranged and reduced in the form*

$$V_{W_i}[k+1] = \max\{V_{W_i}[k] + \sigma_{W_i}[k] + (1-|\mathcal{S}|)(\gamma_i[k] + \sum_{i\in\mathcal{S}} \gamma_i[k]), 0\}, \forall i \in \mathcal{S}, \tag{28}$$

where an outer (lower) bound of $\sigma_{W_i}[k]$ can be estimated as

$$\sigma_{W_i}[k] \geq 2^{|\mathcal{S}|-1}\left[|W||n[k]|\log_2\left(\frac{2\pi e}{\Delta}\right) + \log_2\det(\Sigma_X[k])\right] - \frac{1}{2}\log_2\det(\Sigma_{X_{\tilde{i}^c}}[k]) - (2^{|\mathcal{S}|-3}-2)\sum_{j\in\mathcal{S}\setminus\tilde{i}}\log_2\det(\Sigma_{T_j}[k]). \tag{29}$$

Note that we have shorten the notation $\{i\}$ with $\tilde{i}$ for simplicity.

Proof. Recall that we have defined $V_{W_i}[k] = \sum_{i\in W} V_W[k]$. This can be rewritten into vector form: $\boldsymbol{V}_{W_i}[k] = \boldsymbol{\Omega}_V \boldsymbol{V}_W[k]$, with $\boldsymbol{\Omega}_V$ being the corresponding binary matrix in size $|\mathcal{S}| \times 2^{|\mathcal{S}|}$. Similarly $\boldsymbol{\gamma}_W[k] = \boldsymbol{\Omega}_V^T \boldsymbol{\gamma}_i[k]$. In the kth scheduling period, the scheduler solves an global optimization problem deciding the transmission time ratio, the compression rate, and the amount of data to drop for each sensor node, with coefficients $V_{W_i}[k]$. After solving the program, we update from

$$\boldsymbol{V}_W[k+1] = \max\{\boldsymbol{V}_W[k] - \boldsymbol{\Omega}_V^T\boldsymbol{\gamma}_i[k] + \boldsymbol{\sigma}_W[k], \boldsymbol{0}\}. \tag{30}$$

Possessing $\boldsymbol{V}_W[k+1]$, we can derive $\boldsymbol{V}_{W_i}[k+1]$ for solving the program once again with $\boldsymbol{V}_{W_i}[k+1] = \boldsymbol{\Omega}_V \boldsymbol{V}_W[k+1]$. Then, we can rearrange the virtual queue update equation into:

$$\begin{aligned}\boldsymbol{V}_{W_i}[k+1] &= \boldsymbol{\Omega}_V \max\{\boldsymbol{V}_W[k] - \boldsymbol{\Omega}_V^T\boldsymbol{\gamma}_i[k] + \boldsymbol{\sigma}_W[k], \boldsymbol{0}\} \\ &= \max\{\boldsymbol{V}_{W_i}[k] - \boldsymbol{\Omega}_V\boldsymbol{\Omega}_V^T\boldsymbol{\gamma}_i[k] + \boldsymbol{\Omega}_V\boldsymbol{\sigma}_W[k], \boldsymbol{0}\},\end{aligned} \tag{31}$$

where $\boldsymbol{\Omega}_V\boldsymbol{\Omega}_V^T$ is a $|\mathcal{S}| \times |\mathcal{S}|$ matrix which is time-invariant and can be calculated *a priori* as

$$\boldsymbol{\Omega}_V\boldsymbol{\Omega}_V^T = (|\mathcal{S}|-1)(\boldsymbol{I}_{|\mathcal{S}|} + \boldsymbol{J}_{|\mathcal{S}|}), \tag{32}$$

where $\boldsymbol{I}_{|\mathcal{S}|}$ denotes $|\mathcal{S}| \times |\mathcal{S}|$ identity matrix and $\boldsymbol{J}_{|\mathcal{S}|}$ stands for all-ones matrix in the same size. The computation of term $\boldsymbol{\Omega}_V\boldsymbol{\sigma}_W[k]$ involves matrix-vector multiplication of a $|\mathcal{S}| \times 2^{|\mathcal{S}|}$ matrix and a $2^{|\mathcal{S}|} \times 1$ vector, causing exponential time.

Nevertheless, by applying the properties of logarithm, determinants, and block matrices, we propose an $\mathcal{O}(|\mathcal{S}|^3)$ estimation for its upper bound. To begin, with (31) and (32), we change the transformed virtual queue into scalar form: $\forall i \in \mathcal{S}$,

$$V_{W_i}[k+1] = \max\{V_{W_i}[k] + \sum_{i\in W}\sum_{W\subseteq\mathcal{S}}\sigma_W[k] + (1-|\mathcal{S}|)(\gamma_i[k] + \sum_{i\in\mathcal{S}}\gamma_i[k]), 0\}. \tag{33}$$

We aim to calculate the term with double summation, denoted as $\sigma_{W_i}[k] = \sum_{i\in W}\sum_{W\subseteq\mathcal{S}}\sigma_W[k]$. With (27), employing the fact that $|X[k]| - |X_{W^c}[k]| = |X_W[k]| = |W||n[k]|$ yields:

$$\sigma_{W_i}[k] \approx 2^{|\mathcal{S}|-1}\left[|W||n[k]|\log_2\left(\frac{2\pi e}{\Delta}\right) + \log_2\det(\Sigma_X[k])\right] - \frac{1}{2}\sum_{i\in W}\sum_{W\subseteq\mathcal{S}}\log_2\det(\Sigma_{X_{W^c}}[k]). \tag{34}$$

Obviously, only term $1/2\sum_{i\in W}\sum_{W\subseteq\mathcal{S}}\log_2\det(\Sigma_{X_{W^c}}[k])$ may vary for different i, which can be further rearranged as:

$$\frac{1}{2}\log_2\prod_{W\subseteq\mathcal{S}}\det(\Sigma_{X_{W^c\setminus\tilde{i}}}[k]). \tag{35}$$

From Hadamard-Fischer's inequality, by reasonably assuming $\det(\Sigma_{X_\emptyset}[k]) = 1$ and letting $M = 2^{|\mathcal{S}|-2} - 1$, we have

$$\prod_{W\subseteq\mathcal{S}}\det(\Sigma_{X_{W^c\setminus\tilde{i}}}[k]) \leq \det(\Sigma_{X_{\tilde{i}^c}}[k])\det(\bigoplus_{j\in\mathcal{S}\setminus\tilde{i}}\Sigma_{T_j}[k]^M), \tag{36}$$

and with little rearrangements we can finally come to (29). $\square$

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 1991.

[2] L. Ying, R. Srikant, and G. E. Dullerud, "Multi-user scheduling in wireless networks with qos constraints," in *2006 IEEE International Symposium on Information Theory*, 2006, pp. 2759–2763.

[3] J. Tang and X. Zhang, "Qos-driven adaptive power and rate allocation for multichannel communications in mobile wireless networks," in *2006 IEEE International Symposium on Information Theory*, 2006, pp. 2516–2520.

[4] C. P. S. Thejaswi and V. Sharma, "Adaptive joint source–channel rate allocation policies for delay sensitive applications over fading channels," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5587–5598, 2006.

[5] N. Shlezinger, S. Salamatian, Y. C. Eldar, and M. Médard, "Joint sampling and recovery of correlated sources," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 385–389.

[6] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: maximizing information while minimizing communication cost," in *2006 5th International Conference on Information Processing in Sensor Networks*, 2006, pp. 2–10.

[7] H.-Y. Hsieh, C.-H. Chang, and W.-C. Liao, "Not every bit counts: Data-centric resource allocation for correlated data gathering in machine-to-machine wireless networks," *ACM Transactions on Sensor Networks*, vol. 11, no. 2, pp. 38:1–38:33, Mar 2015.

[8] H.-Y. Hsieh, T.-C. Juan, Y.-D. Tsai, and H.-C. Huang, "Minimizing radio resource usage for machine-to-machine communications through data-centric clustering," *IEEE Transactions on Mobile Computing*, vol. 15, no. 12, pp. 3072–3086, Dec 2016.

[9] H.-Y. Hsieh and C.-Y. Su, "Data-centric scheduling for minimizing queue length in wireless machine-to-machine networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[10] E. N. Ciftcioglu, A. Yener, and M. J. Neely, "Maximizing quality of information from multiple sensor devices: The exploration vs exploitation tradeoff," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 883–894, 2013.

[11] M. Iqbal, T. Aziz, M. Adnan, and M. Abrar, "Relay assisted slepian-wolf compression to exploit temporal redundancy," in *4th International Conference on New Trends in Information Science and Service Science*, 2010, pp. 390–392.

[12] A. Dabirmoghaddam, M. Ghaderi, and C. Williamson, "Cluster-based correlated data gathering in wireless sensor networks," in *2010 IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, 2010, pp. 163–171.

[13] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: Theory and applications for wireless sensor networks," *Computer Networks*, vol. 45, no. 3, pp. 245 – 259, 2004.

[14] C. Yang, X. Zhang, C. Zhong, C. Liu, J. Pei, K. Ramamohanarao, and J. Chen, "A spatiotemporal compression based approach for efficient big data processing on cloud," *J. Comput. Syst. Sci.*, vol. 80, no. 8, p. 1563–1583, Dec 2014.

[15] L. A. Villas, A. Boukerche, D. L. Guidoni, H. A. de Oliveira, R. B. de Araujo, and A. A. Loureiro, "An energy-aware spatio-temporal correlation mechanism to perform efficient data collection in wireless sensor networks," *Computer Communications*, vol. 36, no. 9, pp. 1054 – 1066, 2013.

[16] R. Puri, A. Majumdar, P. Ishwar, and K. Ramchandran, "Distributed video coding in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 94–106, 2006.

[17] T. Wang, A. Vosoughi, W. Heinzelman, and A. Seyedi, "Maximizing gathered samples in wireless sensor networks with slepian-wolf coding," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 751–761, 2012.

[18] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *IEEE INFOCOM 2004*, vol. 4, 2004, pp. 2571–2582.

[19] G. S. Paschos, M. Leconte, and A. Destounis, "Routing with blinkers: Online throughput maximization without queue length information," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 1436–1440.

[20] R. Talak, I. Kadota, S. Karaman, and E. Modiano, "Scheduling policies for age minimization in wireless networks with unknown channel state," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2564–2568.

[21] S. Shadbakht and B. Hassibi, "On the entropy region of gaussian random variables," *CoRR*, vol. abs/1112.0061, 2011. [Online]. Available: http://arxiv.org/abs/1112.0061

[22] T. M. Cover, "A proof of the data compression theorem of slepian and wolf for ergodic sources (corresp.)," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 226–228, March 1975.

[23] R. G. Gallager, *Information Theory and Reliable Communication*. USA: Wiley, 1968.

[24] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 89–103, Jan 2005.

[25] H. P. Benson, "Global maximization of a generalized concave multiplicative function," *J. Optim. Theory Appl.*, vol. 137, no. 1, p. 105–120, Apr 2008.