OUTformation: Distributed Data-Gathering with Feedback under Communication Delay Constraints

SooJean Han[1*], Michelle Effros[1], Richard M. Murray[1]

Abstract— **Towards the informed design of large-scale distributed data-gathering architectures under real-world assumptions such as nonzero communication delays and unknown environment dynamics, this paper considers the effects of allowing feedback communication from the central processor to external sensors. Using simple but representative state-estimation examples, we investigate fundamental tradeoffs between the mean-absolute error (MAE) of the central processor's estimate of the environment state, and the total power cost per sensor, comparing more conventional architectures without feedback, here called *INformation*, to a new strategy using broadcast feedback that we call *OUTformation*). These performance tradeoffs are considered under different architecture parameters and varying levels of environmental variation. The primary advantage of enabling this type of feedback is that each sensor's understanding of the central processor's estimate improves, which helps each sensor determine when to transmit and what parts of its current observations to send. Our results include theoretical bounds under simplifying assumptions, and numerical studies which show that the properties derived in theory hold even when some of the simplifying assumptions are removed.**

I. Introduction

The problem of data-gathering and state-estimation for large-scale cyberphysical networks is important in many applications, such as mapping of a foreign terrain [1], [2], target-tracking [3], and vehicle traffic forecasting [4], [5]. Many distributed data-gathering architectures are built so that multiple lower-level external sensors on the network's edge transmit data to the higher-level processor at the network's center. Variations to this basic architecture have been proposed with the aim of reducing transmissions and minimizing computation, including structural changes to the communication among the sensors (e.g., hierarchical clustering [6], [7]); see [8], [9] for surveys. Other designs have challenged the common assumption that data should flow only in a single direction from the sensors to the central processor, and allow *feedback communication* from the central processor to the sensors. Under simplifying assumptions such as zero communication delays and/or knowledge of the environment dynamics, these feedback architectures have been shown to be just as effective as architectures without feedback at reducing transmissions and computation [10], [11].

In this paper, we show that architectures with feedback perform better distributed data-gathering than those without when two realistic settings are considered: nonzero communication delays and limited knowledge of the environment dynamics. The impact of communication delays is important because delays cause outdated data, which then degrades the performance of distributed state estimation in large-scale networks. This paper discusses when the benefits of feedback in distributed data-gathering exceed the costs, but at a high-level, including feedback provides at least the following two advantages. First, by informing the sensors about what it and isn't known at the central processor, feedback reduces transmission of redundant data. Second, enabling each sensor to transmit less data reduces network traffic and decreases communication delays; this consequently reduces estimation errors.

The contributions of our paper are as follows. Because the type of feedback we consider broadcasts data "out" from the central processor to the sensors, we refer to this architecture as the *OUTformation architecture* for data-gathering. We call the traditional strategy, in which all communication carries data from the sensors to the central processor, the *INformation architecture*. We use a modular representation of simple INformation and OUTformation architectures and take a first step towards characterizing their relative performances with respect to two metrics: 1) the mean-absolute error (MAE) of the central processor's estimate of the environment state, and 2) the power costs of each sensor. In order to maintain tractability, our theoretical analyses tackle simple case studies, each designed to demonstrate the fundamental tradeoff between MAE and power under different environmental statistics and parameter design choices, which is useful in guiding architecture choice to match an application's power constraints and accuracy requirements. We show that the main advantage of feedback is that it enables each sensor to better estimate the central processor's estimate, which guides each sensor towards decisions about when and what to transmit. In order to investigate the sensitivity of our theoretical results to the assumptions employed in their derivations, we include numerical simulations for scenarios that violate those assumptions, and find that they are still consistent with our theory, suggesting that the benefits of the OUTformation architecture extend beyond the framework treated rigorously by our analyses.

II. INformation and OUTformation

A. Problem Setup

We consider an environment represented by an n-dimensional vector $\mathbf{x}$. The environment evolves as a random

[1]SooJean Han, Michelle Effros, and Richard M. Murray are with the Division of Engineering and Applied Science at California Institute of Technology, Pasadena CA.

[*]Corresponding author. Email: soojean@caltech.edu

This paper is based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1745301.

walk, given by

$$\mathbf{x}((c+1)\Delta t) = \mathbf{x}(c\Delta t) + \mathbf{d}(c\Delta t) \tag{1}$$

Here, $c \in \mathbb{N}$ is a time-index, $\Delta t \in \mathbb{R}^+$ is the length of the environment's transition period, and the random vector step-size $\mathbf{d}(c\Delta t)$ is distributed in the following way. At each time $c\Delta t$, a random subset $\mathcal{N}(c\Delta t) \subseteq \{1, \cdots, n\}$ of components is chosen to take nonzero values, with the remaining components zero. For each $i \in \mathcal{N}(c\Delta t)$, the nonzero value $d_i(c\Delta t)$ is uniformly-distributed in $[\underline{d}, \overline{d}]$ with probability $1/2$, otherwise uniformly-distributed in $[-\overline{d}, -\underline{d}]$, for some constant values $0 < \underline{d} < \overline{d}$. The number of components $|\mathcal{N}(c\Delta t)|$ is either fixed constant or randomly-chosen; in this paper, we provide examples of both. For simplicity of expression, every interval $[c\Delta t, (c+1)\Delta t)$, is referred to as the *environment interval* c for $c \in \mathbb{N}$.

We focus on data-gathering architectures where a single central processor communicates with $M \in \mathbb{N}$ independent external sensors. Communications in both uplink and downlink channels are noiseless and, for simplicity, we assume transmission of real numbers. The environment (1) is "unknown" to the central processor and all the sensors in the sense that 1) the parameters $\Delta t, \underline{d}, \overline{d}$ and 2) the distribution of the choice of components in $\mathcal{N}(c\Delta t)$ for each c are unknown. The two different types of architectures we consider are described fully in the following Section II-B.

Each sensor $j \in \{1, \cdots, M\}$ operates on a limited power source (e.g., batteries) and samples $m_j \in \mathbb{N}$ (with $m_j < n$) dimensions of the dimension $\mathbf{x}$ at time intervals of length τ with a constant time-shift of $b_j \geq 0$. We refer to τ as the *(sensor) sampling period* and b_j the *offset* of sensor j. All sampled measurements are perturbed by additive, white Gaussian noise. The resulting measurements are

$$\mathbf{y}_j(a\tau + b_j) = C_j \mathbf{x}(a\tau + b_j) + \mathbf{w}_j(a\tau + b_j). \tag{2}$$

Here, $a \in \mathbb{N}$, and $C_j \in \{0,1\}^{n \times n}$ is a diagonal matrix such that m_j of the diagonal entries are 1, and $C_j(i,i) = 1$ if and only if dimension i of $\mathbf{x}$ is visible to sensor j. Random noise $\mathbf{w}_j(t) \sim$ is zero in all dimensions not visible to sensor j, and the remaining entries are drawn i.i.d. from $\mathcal{N}(\mathbf{0}, \sigma^2)$ for some $\sigma > 0$. The sensors are configured to sample the environment asynchronously from each other via the design of $\{\tau, b_j : j \in \{1, \cdots, M\}$; this is to avoid collision scenarios where different sensors communicate simultaneously with the central processor.

Definition 1 (Observed Dimensions)**.** Define $\mathcal{C}_j \triangleq \{i \in \{1, \cdots, n\} : C_j(i,i) = 1\}$ to be the set of sensor j's *observed dimensions*. Then $\mathcal{S}_j \triangleq \mathcal{C}_j \cap (\cup_{j' \neq j} \mathcal{C}_{j'})$ are the *shared dimensions* of sensor j, while $\mathcal{U}_j \triangleq \mathcal{S}_j^c$ are its *unshared dimensions*. We also define $\mathcal{S} \triangleq \cup_j \mathcal{S}_j$.

B. Modular Framework

In each data-gathering architecture, the central processor uses the transmissions of each sensor to construct an estimate $\hat{\mathbf{x}}(t)$ of the environment state $\mathbf{x}(t)$. Each sensor j uses its own estimate $\hat{\mathbf{x}}_j(t) \triangleq (\hat{x}_{ji}(t) : i \in \{1, \cdots, n\})$ of $\hat{\mathbf{x}}(t)$ to determine which components $y_{ji}(t), i \in \mathcal{C}_j$, of $\mathbf{y}_j(t)$ to transmit, and when to transmit them.

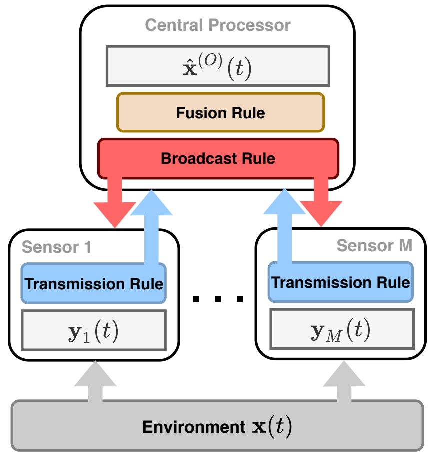

Fig. 1: A modular representation of an OUTformation architecture for distributed state estimation under unknown environments and communication delay constraints, using a single central processor and M external sensors. For each sensor, the transmission rule (Definition 4) is in blue; for the central processor, the fusion rule is in orange and the broadcast rule (Definition 3) is in red. INformation architectures are depicted by removing the red broadcast arrows, and denoting $\hat{\mathbf{x}}^{(O)}(t)$ with $\hat{\mathbf{x}}^{(I)}(t)$ instead.

The presence or absence of *broadcast feedback* (which pushes data "outward" from the central processors to the sensors) changes the central processor's estimate $\hat{\mathbf{x}}(t)$, as captured in the INformation and OUTformation models defined below. A modular representation of the distributed data-gathering OUTformation architecture, which we use throughout the paper, is illustrated in Figure 1.

Definition 2 (INformation and OUTformation)**.** Under an *INformation architecture*, sensor j's estimate $\hat{\mathbf{x}}_j^{(I)}(t)$ of the central processor's estimate $\hat{\mathbf{x}}^{(I)}(t)$ of the true environment state $\mathbf{x}(t)$ contains, in each dimension, its own most recent transmission to the central processor. Under an *OUTformation architecture*, $\hat{\mathbf{x}}_j^{(I)}(t)$ of the central processor's estimate $\hat{\mathbf{x}}^{(O)}(t)$ contains, in each dimension, the more recent of either its previous transmission or the latest broadcast received from the central processor. Here, all estimates are assigned the superscript (I) for INformation architectures, and (O) for OUTformation architectures.

Definition 3 (Broadcast Rule)**.** In this paper, we primarily restrict our attention to scenarios in which only shared dimensions are involved in the central processor's OUTformation transmissions, as described in the following *broadcast rule* (red in Figure 1). For every dimension $i \in \{1, \cdots, n\}$ that is a shared dimension for at least two or more sensors in the network, if $\hat{x}_i^{(O)}(t) \neq \hat{x}_i^{(O)}(s)$, where $s < t$ is the time of the central processor's most recent broadcast, $\hat{x}_i^{(O)}(t)$ is broadcast at time t.

Remark 1. We emphasize that broadcasts are made only for shared dimensions due to the environment dynamics (1) being unknown. However, broadcasting unshared dimensions is beneficial when partial knowledge of the environment is allowed, especially regarding any dependencies among the different dimensions of the environment state $\mathbf{x}$. For concreteness, we carry our analysis throughout the paper under our original setting of unknown environments and broadcast rule Definition 3. An example demonstrating the effectiveness of feedback for data-gathering architectures in partially-known environments is discussed in Section IV-B.

Definition 4 (Transmission Rule)**.** For each architecture $\chi \in \{I, O\}$ and each time $t \triangleq a\tau + b_j$ for some $a \in \mathbb{N}$, sensor $j \in \{1, \cdots, M\}$ uses a *transmission rule* (blue in Figure 1) to check whether the absolute difference between its observation $y_{ji}(t)$ and its estimate $\hat{x}^{(\chi)}_{j,i}(t)$ of the central processor's estimate $\hat{x}^{(\chi)}_i(t)$ of dimension $i \in \{1, \cdots, n\}$ exceeds a certain user-chosen threshold $\epsilon > 0$. Under an INformation architecture, this means checking:

$$\mathbb{1}\left\{|y_{ji}(s_{ji}) - y_{ji}(t)| \geq \epsilon\right\} \tag{3}$$

while under an OUTformation architecture:

$$\begin{cases} \mathbb{1}\{|\hat{x}^{(O)}_i(t) - y_{ji}(t)| \geq \epsilon\} & \text{if } s^* \geq s_{ji} \\ \mathbb{1}\left\{|y_{ji}(s_{ji}) - y_{ji}(t)| \geq \epsilon\right\} & \text{else} \end{cases} \tag{4}$$

Here, $s_{ji} < t$ is the time of sensor j's previous transmission of dimension i, and $s^* < t$ is the time the broadcast value $\hat{x}^{(O)}_i(t)$ was updated. If either (3) or (4) yields 1, depending on the architecture, sensor j transmits $y_{ji}(t)$ to the central processor.

Definition 5 (Fusion Rule)**.** For each $\chi \in \{I, O\}$, the central processor implements the *fusion rule* by component-replacement: if a measurement $y_{ji}(t)$ from sensor j is received at time t, then the estimate is updated as $\hat{x}^{(\chi)}_i(t) = y_{ji}(t)$ for each $i \in \mathcal{C}_j$.

Definition 6. The two types of architectures we consider are specified below:

1) *Absolute-Difference INformation* architecture IN(ϵ): each sensor j implements transmission rule (3) with fixed threshold $\epsilon > 0$. The central processor uses the fusion rule from Definition 5.
2) *OUTformation* architecture OUT(ϵ): each sensor j implements transmission rule (4) with fixed threshold $\epsilon > 0$. The central processor uses the fusion rule from Definition 5 and the broadcast rule from Definition 3.

Remark 2. We emphasize that our goal is not to model any particular distributed data-gathering architecture in fine detail but to investigate the differences between architectures with and without feedback from the central processor to the sensors. Thus, while functionalities (e.g., optimized random backoff times [12], consensus-based fusion [13], change-detection [14]) that are more complex than those described above are possible, we focus specifically on functionalities directly impacted by the presence or absence of feedback.

Definition 7 (Rates per Component)**.** Let the *communication delay* incurred per component be defined as $\Delta t^{(u)} \in \mathbb{R}^+$ for uplink transmission, and $\Delta t^{(d)} \in \mathbb{R}^+$ for downlink broadcast. For sensor j under architecture $\chi \in \{I, O\}$, let $U^{(\chi)}_{ji}(s:t)$ and $D^{(\chi)}_{ji}(s:t) \in \mathbb{N}$ be the cumulative *number of transmissions* and *number of broadcasts received*, respectively, for dimension $i \in \{1, \cdots, n\}$ over the interval of time $[s, t)$. Moreover, define the *rate of power* sensor j expends per dimension i to be $P_U \in \mathbb{R}^+$ for transmission and $P_D \in \mathbb{R}^+$ for receipt. Since broadcasting from the more powerful central processor is faster than the battery-powered sensors, we take $P_U \geq P_D$ and $\Delta t^{(u)} \geq \Delta t^{(d)}$.

Definition 8 (Performance Metrics)**.** The *mean-absolute error (MAE)* of the environment state vector over some experiment duration $[0, T)$, where $T > 0$, under architecture $\chi \in \{I, O\}$ is given by

$$\int_0^T \sum_{i=1}^{n} |x_i(t) - \hat{x}^{(\chi)}_i(t)| dt \tag{5}$$

The *total amount of power cost per sensor* j under architecture χ is given by:

$$P_U \sum_{i \in \mathcal{C}_j} U^{(\chi)}_{ji}(0:T) + P_D \sum_{i \in \{1, \cdots, n\}} D^{(\chi)}_{ji}(0:T) \tag{6}$$

where the power cost rates P_U,P_D are from Definition 7. Under the broadcast rule from Definition 3, the second term in (6) sums over $i \in \mathcal{S}$ instead of $i \in \{1, \cdots, n\}$, where $\mathcal{S}$ is from Definition 1.

III. Analysis

We now theoretically characterize the tradeoff space between the MAE and sensor power cost metrics (see Definition 8) for the three architectures in Definition 6. In order to introduce theory to analyze the fundamental places where OUT(ϵ) can demonstrate improvements over IN(ϵ), we use the following simplifications of the original setup from Section II-A in this section only.

Setting 1 (Analysis Assumptions)**.** The data-gathering architectures from Definition 6 operate with $M = 2$ sensors such that $\Delta t^{(u)} + \Delta t^{(d)} < \tau/2$, offsets $b_1 = 0$ and $b_2 = \tau/2 \in \mathbb{N}$. For each environment, we focus on a specific interval $[t_0, t_f)$ of communications such that $\hat{x}_{ji}(t_0) \triangleq \hat{x}^{(I)}_{ji}(t_0) = \hat{x}^{(O)}_{ji}(t_0)$ equals $y_{ji}(s)$, where $s < t_0$ is the time of the last transmission of component i from sensor j.

A. Reduced Power for Shared Components

Under the broadcast rule of Definition 3, the transmission rule is satisfied the same way under IN(ϵ) or OUT(ϵ), and so $\hat{x}_{ji}(t) = y_{ji}(t)$ for each $i \in \mathcal{U}_j$. Because both IN(ϵ) and OUT(ϵ) have the same uplink communications for unshared dimensions, the total power costs contributed by a sensor's unshared dimensions are equivalent under either architecture.

The more interesting comparison arises for shared dimensions. To demonstrate the effect of shared dimensions without additional delays incurred by other components, we consider the following one-dimensional environment.

Setting 2 (One-Dimensional Environment)**.** In addition to Setting 1, we consider environment dynamics (1) where $n=1$, i.e., $\mathcal{N}(c\Delta t)=\{1\}$. All vector notation $\mathbf{x},\mathbf{y},\mathbf{d}$ are written without boldface as x,y,d, and all additional subscripts of i are removed from the original notation (e.g., $y_{ji}(t),\equiv y_j(t)$). The scalar random step-size $d(c\Delta t)$ is nonzero at each $c\in\mathbb{N}$ and distributed according to the description beneath (1). We choose the specific interval $[t_0,t_f)$ to be environment interval c (i.e., $t_0\triangleq c\Delta t$, $t_f\triangleq(c+1)\Delta t$). Although Δt is unknown, we assume a choice of $\tau<\Delta t$ such that $\Delta t/\tau\in\mathbb{N}$, since the case $\tau\geq\Delta t$ degrades the performance of both architectures and makes for less clear comparisons.

Theorem 1 (Power from Shared Components)**.** Consider the interval $[t_0,t_f)$, architecture assumptions, and environment from Setting 2. If $\epsilon=z^*\sigma$, then OUT(ϵ) consumes P_U-P_D less power than IN(ϵ) over $[t_0,t_f)$ with probability $p(z^*)^{2(\Delta t/\tau-1)}$. Here, σ is defined in (2), z^* is the z-score for a standard normal distribution, and $p(z^*)$ is its corresponding confidence interval probability.

Proof. Under Setting 2, sensor 2 does not transmit at time $t_0+(\tau/2)$ only if the sensor observations do not satisfy the transmission rule (i.e., $|y_1(t_0)-y_2(t_0)|<\epsilon$). For the remainder of the $(\Delta t/\tau)-1$ samples in $[t_0,t_f)$ for each sensor j, the transmission rule must not be satisfied with $\hat{x}_j(t)=y_j(t)$. By (2), each occurs with probability $\mathbb{P}(|W_1-W_2|<\epsilon)$ where $W_1,W_2\sim\mathcal{N}(0,\sigma^2)$ are independent and σ is from (2). The result follows from standard confidence intervals and the fact that P_U-P_D is the difference between transmitting and broadcasting an extra component (see Definition 7). $\square$

Remark 3 (Generalization of Theorem 1)**.** Theorem 1 suggests that for good parameter choice of ϵ with respect to σ, OUT(ϵ) is able to spend less power on average compared to IN(ϵ) from shared dimensions. More general implications of Theorem 1 are shown with Figure 2, which plots the communications over time in the original setup of Section II with the assumptions of Settings 1 and 2 removed, and when $M=5$ with 12 out of the $n=20$ dimensions shared. Compared to IN(ϵ), OUT(ϵ) has less uplink transmissions overall due to broadcasts updating each sensor's estimate of the central processor's estimate; this is illustrated in Figure 2 by missing or steeper-sloped lines in OUT(ϵ) compared to IN(ϵ). Thus, multiple sensors observing the same shared dimension may decrease their total power cost from shared dimensions at the expense of additional power from a few sensors. We formalize this in the following corollary.

Corollary 1 (Theorem 1 for More Sensors)**.** Consider the setting of Theorem 1, but with a general $M>2$ sensors such that $\Delta t^{(u)}+\Delta t^{(d)}>b_j$ for a subset $m<M$ of the sensors. Then if $\epsilon=z^*\sigma$, then OUT(ϵ) consumes $(M-m)P_U-mP_D$ less power from shared dimensions than IN(ϵ) over $[t_0,t_f)$ with probability $p(z^*)^{\sum_j\lceil((\Delta t-b_j)/\tau)\rceil}$, for notations defined in Theorem 1.

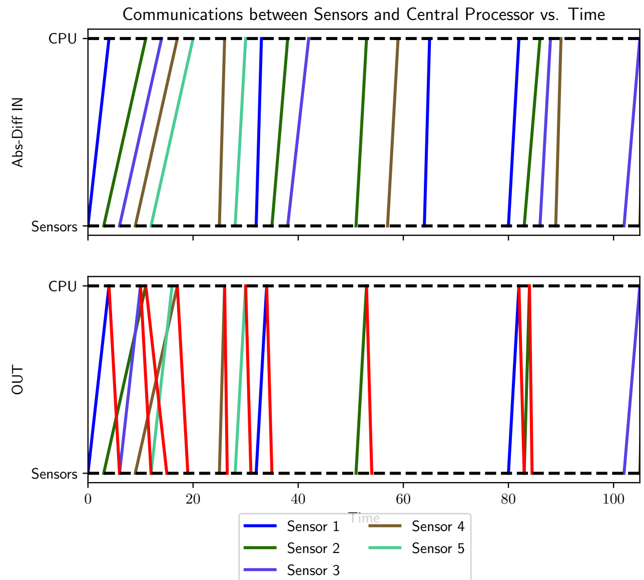

Fig. 2: An illustration of the results of Theorem 1 in the original setup in Section II with the assumptions of Settings 1 and 2 removed. Plots the time-evolution of communications between a central processor and $M=5$ sensors, with 12 out of the $n=20$ dimensions shared, under IN(ϵ) [Top] and OUT(ϵ) [Bottom] from Definition 6. Here, $\Delta t=25$, $\tau=13$, $b_j=3(j-1)$,$\sigma=0.1$, and $\epsilon=2.0$. The uplink transmissions of each sensor are colored according to the shown legend, and central processor broadcasts are in red. The slopes of each line vary because communication delays are proportional to the number of components transmitted/broadcast (see Definition 7).

B. Reduced MSE for Unshared Components

In Section III-A, we argued that the power cost contributed by the unshared dimensions is equivalent under IN(ϵ) and OUT(ϵ). Here, we show that OUT(ϵ) can allow faster delivery of unshared dimensions to the central processor, and discuss conditions for which the MAE contributed by the unshared dimensions is less under OUT(ϵ) than IN(ϵ). We now consider the following two-dimensional environment.

Setting 3 (Two-Dimensional Environment)**.** We consider environment dynamics (1) where $n=2$ and there occurs a change in exactly one dimension $i\in\{1,2\}$ per each environment interval $c\in\mathbb{N}$, i.e., $|\mathcal{N}(c\Delta t)|=1$ and the element in $\mathcal{N}(c\Delta t)$ is one randomly-chosen i. Sensor 1 observes dimensions 1 and 2, while sensor 2 observes only dimension 2. We choose the specific interval $[t_0,t_f)$ to be $t_0\triangleq(a_1-1)\tau$ and $t_f\triangleq(c+2)\Delta t$ such that $(a_1-1)\tau<a_2\tau+b_2<a_1\tau<(c+2)\Delta t$ and $\Delta t^{(u)}+\Delta t^{(d)}\leq a_1\tau-(a_2\tau+b_2)$ for some $a_1,a_2,c\in\mathbb{N}$. Moreover, shared dimension 2 changes value once by magnitude d_2 at time $c\Delta t\in(t_0,a_2\tau+b_2]$ and remains constant during $(a_2\tau+b_2,t_f]$, and unshared dimension 1 of sensor 1 remains constant during $(t_0,a_2\tau+b_2]$ and changes value once by magnitude d_1 at time $(c+1)\Delta t\in(a_2\tau+b_2,t_f]$. Here, d_1,d_2 are distributed according to the stepsize distribution of (1).

Theorem 2 (MAE from Unshared dimensions)**.** Consider the interval $[t_0,t_f)$, architecture assumptions, and environment from Setting 3. Then OUT(ϵ) incurs on average $2\Delta t^{(u)}\mathbb{E}[|d_1+W_1|-|W_2|]\Phi(\epsilon/\sqrt{2}\sigma)$ less MAE contribution from unshared dimension 1 than IN(ϵ) over $[t_0,t_f)$, where Φ is the cdf of the standard normal distribution, d_1 is as in Setting 3, and $W_1,W_2\sim\mathcal{N}(0,\sigma^2)$ are independent.

Proof. Under IN(ϵ), sensor 1 transmits both dimensions 1 and 2. Under OUT(ϵ) at time $a_2\tau+b_2+\Delta t^{(u)}+\Delta t^{(d)}$, sensor 1 only transmits its unshared dimension 1, because it receives a broadcast update $y_{22}(a_2\tau+b_2)$. This happens when $|y_{22}(a_2\tau+b_2)-y_{12}(a_1\tau)|<\epsilon$, which is equivalent to $|W_{22}-W_{12}|<\epsilon$ for some independent $W_{12},W_{22}\sim\mathcal{N}(0,\sigma^2)$; this occurs with probability $2\Phi(\epsilon/\sqrt{2}\sigma)$. Under IN($\epsilon$), the central processor then receives $y_{11}(a_1\tau)$ with an additional uplink delay of $\Delta t^{(u)}$ compared to OUT(ϵ), and the difference in MAE of dimension 1 is

$$\Delta t^{(u)}(|\hat{x}_{11}(t_0)-x_1(c\Delta t)|-|y_{11}(a_1\tau)-x_1(c\Delta t)|). \quad (7)$$

By Setting 3 and (2), we have $|\hat{x}_{11}(t_0)-x_1(c\Delta t)|=|d_1+W_1|$ and $|y_{11}(a_1\tau)-x_1(c\Delta t)|=|W_2|$ for some independent $W_1,W_2\sim\mathcal{N}(0,\sigma^2)$ which are also independent of W_{11},W_{22}. The result follows. $\square$

Theorem 2 suggests that for good parameter choice of ϵ with respect to σ, OUT(ϵ) yields less MAE on average from the unshared dimensions than IN (ϵ). This is because, by transmission rule (4), OUT(ϵ) transmits less components on average among the two architectures, and so it also incurs the least uplink delays. Because sensors are able to alleviate the burden of transmitting shared dimensions from other sensors, the central processor under OUT(ϵ) detects changes in the unshared dimensions of the environment state more quickly than IN(ϵ), consequently decreasing MAE.

Remark 4 (Generalization of Theorem 2)**.** More general implications of Theorem 2 are shown with Figure 3, which plots the evolution of the estimates $\hat{\mathbf{x}}_i^{(\chi)}(t),\chi\in\{I,O\}$ and the true state $x_i(t)$ over time for three dimensions $i\in\{2,6,10\}$; the original setup of Section II is used with the assumptions of Settings 1 and 3 removed, and when $M=3$ with 4 out of the $n=12$ dimensions shared. Here, dimensions 2 and 10 are unshared whereas 6 is shared by two sensors. Indeed, we observe that the central processor under OUT(ϵ) detects changes in the true environment state more quickly for the unshared dimensions. Moreover, we observe that the tracking performance of the shared dimension is the same on average for both architectures. This is because shared dimensions are transmitted less often on average under OUT(ϵ) than under IN(ϵ) (recall Theorem 1), and silence from a sensor means that its measurement is within ϵ difference of the central processor's estimate, which is an insignificant contribution to the MAE for a good choice of ϵ with respect to the minimum step-size $\underline{d}$.

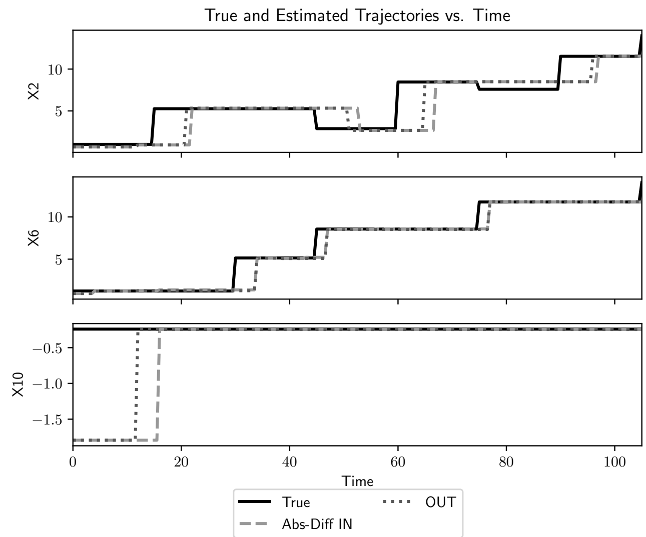

Fig. 3: An illustration of the results of Theorem 2 in the original setup in Section II with the assumptions of Settings 1 and 3 removed. Plots the evolution of the estimates $\hat{\mathbf{x}}_i^{(\chi)}(t),\chi\in\{I,O\}$ and the true state $x_i(t)$ over time for three dimensions $i\in\{2,6,10\}$. Here, $M=3$ with 4 out of the $n=12$ dimensions shared, $\Delta t=15$, $\tau=15$, $b_j=4(j-1)$,$\sigma=0.1$, and $\epsilon=2.0$. We see that $\hat{\mathbf{x}}^{(O)}(t)$ tracks $\mathbf{x}(t)$ more accurately than $\hat{\mathbf{x}}^{(I)}(t)$ dimensions 2 and 10, which are unshared, and just as accurately for the dimension 6, which is shared by two sensors.

IV. More General Insights

A. Numerical Demonstration

In this section, we supplement the insights derived from the previous Section III by empirically plotting the tradeoff space between the performance metrics of Definition 8 in the setup of Figure 2, which does not operate under any of the assumptions of Settings 1, 2, and 3. The environment (1) is simulated with $[\underline{d},\overline{d}]=[2.0,4.0]$, $\Delta t=10.0$, and has dimension $n=20$; the number of components $|\mathcal{N}(c\Delta t)|$ which are allowed to change at each $c\Delta t$ is a random number between 5 and 10. Both architectures IN(ϵ) and OUT(ϵ) are designed with threshold $\epsilon<1.0$ and $M=5$ sensors with white noise variance $\sigma=0.2$; sensor 1 observes dimensions 9 to 12, sensor 2 observes dimensions 5 to 12, sensor 3 observes dimensions 9 to 16, sensor 4 observes dimensions 1 to 8, and sensor 5 observes dimensions 13 to 20. We choose communication delays $\Delta t^{(d)}=0.5$ and $\Delta t^{(u)}=1.0$ (so that $\Delta t^{(d)}<\Delta t^{(u)}$) and power cost rates $P_D=1.0$ and $P_U=2.0$ (so that $P_D<P_U$).

To the above setup, we also make the following modifications. For OUT(ϵ), we additionally vary over $p^{(b)}\in[0,1]$, a constant probability in which the central processor makes a broadcast upon receiving a transmission. For significance, the fusion rule we use is not the component-replacement rule described in Definition 5; rather, the central processor performs fusion by averaging over a finite, fixed horizon of the most recent sensor observations it received to construct an estimate $\hat{x}_i(t)$ for each component $i\in\{1,\cdots,n\}$.

The results are demonstrated in Figure 4 for sensors 3 and 4 of the five sensors, both of which have 4 unshared dimensions and 4 shared dimensions, and the metrics are computed over 20 Monte-Carlo trials. With a larger number $M \geq 2$ of sensors, one may be inclined to believe that IN(ϵ) obtains a lower MAE under the finite-horizon averaging fusion rule because it transmits the largest number of independent, redundant noisy observations, and the law of large numbers would take effect. However, when communication delays and offset sensor sample transmissions are involved, a larger number of transmissions will take a longer delay to be received by the central processor, causing an increase in MAE. Thus, the two effects balance out, and we observe exactly the same insights we've derived from Remarks 3 and 4 (and Figures 2 and 3). Namely, we observe Remark 3 in Figure 4 from the way sensor 3 decreases its total power under OUT(ϵ) than IN(ϵ), while sensor 4 increases; multiple sensors observing the same shared dimension may decrease their power cost from shared dimensions at the expense of additional power from a few other sensors, and extra power comes from the unshared dimensions. Moreover, one sensor expends more power for both transmitting and receiving broadcasts, while the sensor expends mostly downlink power for receiving broadcast updates; in Figure 4, this tradeoff is shown explicitly: increasing downlink power (in gold) allows for decreasing uplink power (in green) for both sensors. We observe Remark 4 in Figure 4 from the way both sensors 3 and 4 attain less average MAE under OUT(ϵ) than IN(ϵ). These insights occur for $p^{(b)} = 1$, and Figure 4 suggests that a balance between total power cost and MAE may be obtained not only through a good choice of ϵ with respect to σ, but also by choosing $p^{(b)}$ appropriately.

B. Extension to Partially-Known Environments

We investigated the benefits of OUTformation in distributed data-gathering under unknown environment statistics. However, continuing from Remark 1, we argue that OUTformation can attain better performance than INformation even with some partial knowledge of the environment assumed. In this section, we demonstrate the potential power cost advantage of OUT(ϵ) over IN(ϵ) with the following simple example.

The environment evolves according to the following two-dimensional stochastic dynamics:

$$\mathbf{x}(c\Delta t) = \begin{cases} [d_1(c\Delta t), 0]^\top & \text{w.p. } p \\ [0, d_2(c\Delta t)]^\top & \text{w.p. } 1-p \end{cases} \quad (8)$$

Here, c and Δt are defined in (1), $p \in (0,1)$ is constant, and the random step-sizes $\{d_1(c\Delta t), \cdots, d_n(c\Delta t)\}$ are distributed as in (1). "Partial knowledge" about (8) refers to the following two conditions about the knowledge of the central processor and all the sensors. First, Δt, p, and random step-size limits $\{\underline{d}, \overline{d}\}$ (defined in (1)) are unknown. Second, the fact that only one component of the state can take a nonzero value at each time is known. The architectures IN(ϵ) and OUT(ϵ) (Definition 6) are built according to Setting 1 such that sensor $j \in \{1, 2\}$ observes dimension j and $\Delta t^{(u)} + 2\Delta t^{(b)} \leq (\tau/2)$

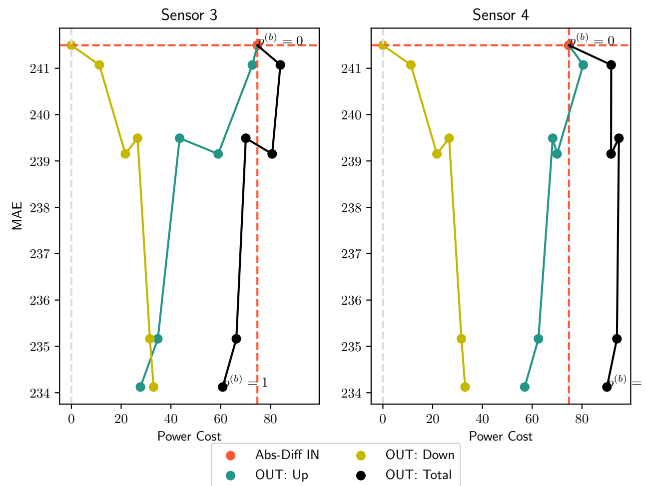

Fig. 4: Plots of the MAE against power cost for two sensors 3 and 4, averaged over 20 Monte-Carlo simulations of IN(ϵ) and OUT(ϵ) over the setup of Figure 2 with the finite-horizon averaging fusion rule (see Section IV-A). The orange dot is for IN(ϵ), with dashed lines to emphasize its visibility. For OUT(ϵ), a constant broadcast probability $p^{(b)}$ is ranged over $\{0, 0.2, \cdots, 1\}$, and we make separate curves for the uplink (green), downlink (gold), and the total (black) power costs. We see a combination of effects from Figures 2 and 3: under OUT(ϵ), both sensors are able to attain smaller MAE (contributed by the unshared components) than IN(ϵ) by having a subset of the sensors alleviate the power cost of transmitting shared dimensions from the others. A balance between total power cost and MAE may be obtained by adjusting $p^{(b)}$.

For this setup, there are several key differences from the analysis under unknown environments. First, the broadcast rule (Definition 3) should also transmit unshared dimensions. Second, the central processor's fusion rule should take advantage of the partial knowledge.

Definition 9 (Fusion Rule with Partial Knowledge)**.** Let $\theta > 0$ be a user-chosen threshold parameter, and suppose its estimate $\hat{\mathbf{x}}(t)$ at time t is the all-zero vector with a single nonzero value $\hat{x}_i(t)$ at component i. Suppose it then receives a sensor transmission $z_{i'}(t)$ at time t for component i'. If $i' = i$ and $|z_{i'}(t) - \hat{x}_i(t)| > \theta$, then running-average fusion is performed. If $i' \neq i$ and $|z_{i'}(t)| < \theta$, then the current estimate $\hat{\mathbf{x}}(t)$ is kept as is. If neither of the above two conditions are satisfied, then the estimate is changed to the all-zero vector with a single nonzero value $z_{i'}(t)$ at new component i'. This design of the fusion rule suggests an intuitive choice of $\theta = \epsilon$.

The interval $[t_0, t_f)$ from Setting 1 is chosen to be environment interval c (i.e., $t_0 \triangleq c\Delta t$ and $t_f \triangleq (c+1)\Delta t$) such that there was a state change from $\mathbf{x}((c-1)\Delta t) = [0, d_2]^\top$ to $\mathbf{x}(t_0) = [d_1, 0]^\top$ for some $d_1 \in [\underline{d}, \overline{d}]$. Sensor 1 transmits $y_{11}(t_0)$ at time t_0 with probability

$$\mathbb{P}(|y_{11}(t_0) - \hat{x}_{11}(t_0)| \geq \epsilon)\mathbb{P}(|d_1 + (W_1' - W_1)| \geq \epsilon)$$

for some independent $W_1', W_1 \sim \mathcal{N}(0, \sigma^2)$. Then under either architecture, the central processor receives $y_{11}(t_0)$ at time $t_0 + \Delta t^{(u)}$; it updates its entire estimate vector

$\hat{\mathbf{x}}^{(\chi)}(t_0 + \Delta t^{(u)}) = [y_{11}(t_0), 0]^\top$ according to the fusion rule described in Definition 9, for $\chi \in \{I, O\}$.

The behavior of sensor 2 differs based on the type of architecture. Under $IN(\epsilon)$, sensor 2 transmits $y_{22}(t_0 + b_2)$ to the central processor, and at time $t \triangleq t_0 + b_2 + \Delta t^{(u)}$, $\hat{\mathbf{x}}^{(I)}(t) = [y_{11}(t_0), 0]^\top$ with probability

$$\mathbb{P}(|y_{22}(t_0 + b_2) - \hat{x}_{22}(t_0)| \geq \epsilon) = \mathbb{P}(|W_2' - (d_2 + W_2)| \geq \epsilon)$$

for some independent $W_2', W_2 \sim \mathcal{N}(0, \sigma^2)$. Under $OUT(\epsilon)$, the central processor broadcasts $[y_{11}(t_0), 0]^\top$, which is received by sensor 2 at time $t_0) + \Delta t^{(u)} + 2\Delta t^{(b)}$. With probability

$$\mathbb{P}(|y_{22}(t_0 + b_2) - 0| < \epsilon) = \mathbb{P}(|W_2'| < \epsilon)$$

sensor 2 does not transmit and a reduction in power is observed if $2P_D < P_U$.

All of the probabilities described are large with large $\underline{d}, \overline{d}$ and small σ, and so similar to Theorem 1, we see that a good choice of ϵ with respect to σ allows OUT(ϵ) to experience a decrease in power cost compared to IN(ϵ). However, our example also suggests that leveraging any partial information about the environment allows for a power reduction to arise from unshared dimensions in addition to shared dimensions. Moreover, this advantage comes with the scheduling of sensor offsets $\{b_j\}$ to accommodate longer broadcast delays, since we are now broadcasting both unshared and shared dimensions.

V. Conclusion

We provided a theoretical and numerical characterization of the tradeoff space for architectures with broadcast feedback (OUTformation) and architectures without feedback (INformation) using two performance metrics: the mean-absolute error (MAE) of the central processor's estimate of the environment state and the total power cost per sensor. Our study was motivated towards enabling users to make informed design choices in distributed data-gathering architectures for large-scale network environments under real-world constraints such as nonzero communication delays and limited knowledge of the environment dynamics. We found that on average, under good parameter choices of transmission threshold ϵ with respect to measurement noise standard deviation σ and sensor transmission times $\{\tau_j, b_j : j \in \{1, \cdots, M\}\}$, OUTformation architectures consume less uplink power than INformation, which implies less redundant transmissions. Moreover, under the same conditions, OUTformation has less MAE incurred by the unshared dimensions of the environment than INformation, which implies faster detection of changes in the environment. Although both advantages come at additional downlink power that potentially increases the total power cost, a better tradeoff between the MAE and the total power cost can be attained by varying the broadcast probability $p^{(b)}$. In conclusion, the main advantage of feedback is that each sensor's understanding of the central processor's state estimate improves, and the analysis of these tradeoffs for architectures more complex broadcast, transmission, and fusion rules is a natural subject of future work.

References

[1] D. Tardioli, D. Sicignano, L. Riazuelo, A. Romeo, J. L. Villarroel, and L. Montano, "Robot teams for intervention in confined and structured environments," *J. Field Robotics*, vol. 33, pp. 765–801, 2016.

[2] "DARPA Subterranean (SubT) challenge," https://www.subtchallenge.com/, accessed: 2020-02-20.

[3] S. Bandyopadhyay and S.-J. Chung, "Distributed bayesian filtering using logarithmic opinion pool for dynamic sensor networks," *Automatica*, vol. 97, pp. 7–17, Nov 2018.

[4] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting," in *International Conference on Learning Representations*, 2018.

[5] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, ser. IJCAI'18. AAAI Press, 2018, p. 3634–3640.

[6] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *IEEE INFOCOM 2003.*, vol. 3, 2003, pp. 1713–1723.

[7] O. Younis and S. Fahmy, "Distributed clustering in ad-hoc sensor networks: a hybrid, energy-efficient approach," in *IEEE INFOCOM 2004.*, vol. 1, 2004.

[8] F. Castanedo, "A review of data fusion techniques," *The Scientific World Journal*, vol. 2013, pp. 1–19, Oct 2013.

[9] M. Liggins, C.-Y. Chong, I. Kadar, M. Alford, V. Vannicola, and S. Thomopoulos, "Distributed fusion architectures and algorithms for target tracking," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 95–107, 1997.

[10] B. Chen, W.-A. Zhang, L. Yu, G. Hu, and H. Song, "Distributed fusion estimation with communication bandwidth constraints," *IEEE Transactions on Automatic Control*, vol. 60, no. 5, pp. 1398–1403, 2015.

[11] X. Ge and Q.-L. Han, "Distributed event-triggered H_∞ filtering over sensor networks with communication delays," *Information Sciences*, vol. 291, pp. 128 – 142, 2014.

[12] G. Nutt and D. Bayer, "Performance of CSMA/CD Networks Under Combined Voice and Data Loads," *IEEE Transactions on Communications*, vol. 30, no. 1, pp. 6–11, 1982.

[13] R. Olfati-Saber and J. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Proceedings of the 44th IEEE Conference on Decision and Control*, 2005, pp. 6698–6703.

[14] P. Fearnhead and G. Rigaill, "Changepoint detection in the presence of outliers," *Journal of the American Statistical Association*, vol. 114, no. 525, pp. 169–183, 2019.