An Event-based Diffusion LMS Strategy

Yuan Wang, Wee Peng Tay
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore
Email: ywang037@e.ntu.edu.sg, wptay@ntu.edu.sg

Wuhua Hu
Department of Artificial Intelligence for Applications
SF Technology Co. Ltd
Shenzhen, China
Email: wuhuahu@sf-express.com

***Abstract*—We consider an event-based communication mechanism for diffusion least mean-squares estimation in a sensor network, in which an intermediate estimate from a sensor is communicated to its neighbors only when a triggering criterion is satisfied. We provide a sufficient condition for the mean error stability of our proposed event-based diffusion strategy, and derive an upper bound of its steady-state network mean-square deviation (MSD). Simulations demonstrate that our event-based strategy can achieve similar steady-state network MSD as the adapt-then-combine diffusion strategy but at a significantly lower communication rate.**

***Index Terms*—Diffusion adaptation, energy-efficiency, event-based communication**

I. Introduction

In the era of big data and Internet-of-Things (IoT), ubiquitous smart devices continuously sense the environment and generate large amount of data rapidly. To better address the real-time challenges arising from online inference, optimization and learning, distributed adaptation algorithms have become especially promising and popular compared with traditional centralized solutions. As computation and data storage resources are distributed to every sensor node in the network, information can be processed and fused through local cooperation among neighboring nodes, and thus reducing system latency and improving robustness and scalability. Among various implementations of distributed adaptation solutions [1]–[4], diffusion strategies are particularly advantageous for continuous adaptation using constant step-sizes, thanks to their low complexity, better mean-square deviation (MSD) performance and stability [5]–[9]. Therefore diffusion strategies have attracted a lot of research interests in recent years for both single-task scenarios where nodes share a common parameter of interest [10]–[15], and multi-task networks where parameters of interest differ among nodes or groups of nodes [16]–[20].

In diffusion strategies, each sensor communicates local information to their neighboring sensors in each iteration. However, in IoT networks, devices or nodes usually have limited energy budget and communication bandwidth, which prevent them from frequently exchanging information with neighboring sensors. Several methods to improve energy efficiency in diffusion have been proposed in the literature, the works [21]–[23] aim at reducing the number of neighbors to cooperate with, whereas [24]–[26] reduces the dimension of the local information to be transmitted. These methods either rely on additional optimization procedures, or use auxiliary selection or projection matrices, which require more computation resources to implement.

Unlike time-driven communication where nodes exchange information at every iteration, in event-based communication mechanisms, nodes only trigger communication with neighbors upon occurrence of certain meaningful events. This can significantly reduce energy consumption by avoiding unnecessary information exchange especially when the system has reached steady state. It also allows every node in the network to share the limited bandwidth resource so that channel efficiency is improved. Such mechanisms have been developed for state estimation, filtering, and distributed control over wireless sensor networks [27]–[31], but have not been fully investigated in the context of diffusion adaptation. In [32], the authors proposed a diffusion strategy where entries of the intermediate estimates are quantized into multiple predefined levels, and certain quantized entry of the latest intermediate estimate is communicated if it jumps to a different quantization level from the previous iteration. The performance of this method relies largely on the precision of the selected quantization scheme, which is difficult to optimize for online adaptation where the parameter of interest and environment may change over time.

In this paper, we propose an event-based diffusion strategy to reduce communication among neighboring nodes while preserve the advantages of diffusion strategies. Specifically, each node monitors the difference between the full vector of its current local update and the most recent intermediate estimate transmitted to its neighbors. A communication is triggered only if this difference is sufficiently large. We provide a sufficient condition for the mean error stability of our proposed strategy, and an upper bound of its steady-state network MSD. Simulations demonstrate that our event-based strategy achieves a similar steady-state network MSD as the popular adapt-then-combine (ATC) diffusion strategy but a significantly lower communication rate.

The rest of this paper is organized as follows. In Section II, we introduce the network model, problem formulation and discuss prior works. In Section III, we describe our proposed event-based diffusion LMS strategy and analyze its performance. Simulation results are demonstrated in Section IV, followed by concluding remarks in Sections V.

II. Data Models and Preliminaries

A. Network and Data Model

Consider a network represented by an undirected graph $G = (V, E)$, where $V = \{1, 2, \cdots, N\}$ denotes the set of nodes, and E is the set of edges. Any two nodes are said to be connected if there is an edge between them. The neighborhood of each node k is denoted by $\mathcal{N}_k$ which consists of node k and all the nodes connected with node k. Since the network is assumed to be undirected, if node k is a neighbor of node ℓ, then node ℓ is also a neighbor of node k. Without loss of generality, we assume that the network is connected.

Every node in the network aims to estimate an unknown parameter vector $w^\circ \in \mathbb{R}^{M\times 1}$. At each time instant $i \geq 1$, each node k observes data $\boldsymbol{d}_k(i) \in \mathbb{R}$ and $\boldsymbol{u}_k(i) \in \mathbb{R}^{M\times 1}$, which are related through the following linear regression model:

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_k^\mathsf{T}(i) w^\circ + \boldsymbol{v}_k(i), \tag{1}$$

where $\boldsymbol{v}_k(i)$ is an additive observation noise. We make the following assumptions.

Assumption 1. *The regression process $\{\boldsymbol{u}_{k,i}\}$ is zero-mean, spatially independent and temporally white. The regressor $\boldsymbol{u}_k(i)$ has positive definite covariance matrix $R_{u,k} = \mathbb{E}\left[\boldsymbol{u}_k(i)\boldsymbol{u}_k^\mathsf{T}(i)\right]$.*

Assumption 2. *The noise process $\{\boldsymbol{v}_k(i)\}$ is spatially independent and temporally white. The noise $\boldsymbol{v}_k(i)$ has variance $\sigma_{v,k}^2$, and is assumed to be independent of the regressors $\boldsymbol{u}_\ell(j)$ for all $\{k, \ell, i, j\}$.*

B. ATC Diffusion Strategy

To estimate the parameter w°, the network solves the following least mean-squares (LMS) problem:

$$\min_w \sum_{k=1}^{N} J_k(w), \tag{2}$$

where for each $k \in V$,

$$J_k(w) = \sum_{k\in\mathcal{N}_k} \mathbb{E}\left|\boldsymbol{d}_k(i) - \boldsymbol{u}_k(i)^\mathsf{T} w\right|^2. \tag{3}$$

The ATC diffusion strategy [5], [8] is a distributed optimization procedure that attempts to solve (2) iteratively by performing the following local updates at each node k at each time instant i:

$$\begin{aligned}\boldsymbol{\psi}_k(i) = \boldsymbol{w}_k(i-1)& \\ + \mu_k \boldsymbol{u}_k(i)\left(\boldsymbol{d}_k(i) - \boldsymbol{u}_k(i)^\mathsf{T}\boldsymbol{w}_k(i-1)\right)&, \qquad (4)\\ \boldsymbol{w}_{k,i} = \sum_{\ell\in\mathcal{N}} a_{\ell k}\boldsymbol{\psi}_{\ell,i}&, \qquad (5)\end{aligned}$$

where $\mu_k > 0$ is a chosen step size. The procedure in (4) is referred to as the *adaptation* step and (5) is the *combination* step. The combination weights $\{a_{\ell k}\}$ are non-negative scalars and satisfy:

$$a_{\ell k} \geq 0,\ \sum_{\ell=1}^{N} a_{\ell k} = 1,\ a_{\ell k} = 0, \text{ if } \ell \notin \mathcal{N}_k. \tag{6}$$

Algorithm 1 Event-based ATC Diffusion Strategy (EB-ATC)

for every node k at each time instant i **do**
  Update intermediate estimate $\boldsymbol{\psi}_k(i)$ using (4)
  Compute $\boldsymbol{\epsilon}_k^-(i)$ and $f\left(\boldsymbol{\epsilon}_k^-(i)\right)$.
  if $f\left(\boldsymbol{\epsilon}_k^-(i)\right) > \delta_k(i)$ **then**
    (i) Trigger the communication, broadcast local update $\boldsymbol{\psi}_k(i)$ to every neighbors $\ell \in \mathcal{N}_k$.
    (ii) Mark $\boldsymbol{\gamma}_k(i) = 1$, and update $\overline{\boldsymbol{\psi}}_k(i) = \boldsymbol{\psi}_k(i)$.
  else if $f\left(\boldsymbol{\epsilon}_k^-(i)\right) \leq \delta_k(i)$ **then**
    (i) Keep silent.
    (ii) Mark $\boldsymbol{\gamma}_k(i) = 0$, and update $\overline{\boldsymbol{\psi}}_k(i) = \overline{\boldsymbol{\psi}}_k(i-1)$.
  end if
  Perform diffusion combination
  $\boldsymbol{w}_k(i) = a_{kk}\boldsymbol{\psi}_k(i) + \sum_{\ell\in\mathcal{N}_k\setminus\{k\}} a_{\ell k}\overline{\boldsymbol{\psi}}_\ell(i)$
end for

The local estimates $\boldsymbol{w}_{k,i}$ in the ATC strategy are shown to converge in mean to the true parameter w° if the step sizes μ_k are chosen to be below a particular threshold [5], [8].

III. Event-Based Diffusion

A. Event-based Communication

We consider a modification of the ATC strategy so that the local intermediate estimate $\boldsymbol{\psi}_k(i)$ of each node k is communicated to its neighbors only at certain triggering time instants s_k^n, $n = 1, 2, \ldots$. Let $\overline{\boldsymbol{\psi}}_k(i)$ be the last local intermediate estimate node k transmitted to its neighbors at time instant i, i.e.,

$$\overline{\boldsymbol{\psi}}_k(j) = \boldsymbol{\psi}_k(s_k^n), \text{ for } j \in \left[s_k^n, s_k^{n+1}\right). \tag{7}$$

Let $\boldsymbol{\epsilon}_k^-(i)$ be the *a prior* gap defined as

$$\boldsymbol{\epsilon}_k^-(i) = \boldsymbol{\psi}_k(i) - \overline{\boldsymbol{\psi}}_k(i-1). \tag{8}$$

Let $f\left(\boldsymbol{\epsilon}_k^-(i)\right) = \left\|\boldsymbol{\epsilon}_k^-(i)\right\|_{Y_k}^2$, where Y_k is a positive semi-definite weighting matrix, and $\|P\|_\Sigma^2 = P^T\Sigma P$ for any compatible matrix P.

For each node k, transmission of its local intermediate estimate $\boldsymbol{\psi}_k(i)$ is triggered whenever

$$f\left(\boldsymbol{\epsilon}_k^-(i)\right) > \delta_k(i) > 0, \tag{9}$$

where $\delta_k(i)$ is the threshold adopted by node k at time i.

In this paper, we allow the thresholds to be time-varying. We further assume $\{\delta_k(i)\}$ of each node k are upper bounded, and let

$$\delta_k = \sup\{\delta_k(i) | i > 0\}. \tag{10}$$

In addition, we define binary variables $\{\boldsymbol{\gamma}_k(i)\}$ such that $\boldsymbol{\gamma}_k(i) = 1$ if node k transmits at time instant i, and 0 otherwise. The sequence of triggering time instants $0 \leq s_k^1 \leq s_k^2 \leq \ldots$ can then be defined recursively as

$$s_k^{n+1} = \min\{i \in \mathbb{N} | i > s_k^n, \boldsymbol{\gamma}_k(i) = 1\}. \tag{11}$$

For every node in the network, we apply the event-based adapt-then-combine (EB-ATC) strategy detailed in Algorithm 1. A succinct form of the EB-ATC can be summarized as the following equations,

$$\boldsymbol{\psi}_k(i) = \boldsymbol{w}_k(i-1) + \mu_k \boldsymbol{u}_k(i)\left(\boldsymbol{d}_k(i) - \boldsymbol{u}_k(i)^\mathsf{T}\boldsymbol{w}_k(i-1)\right), \quad (12)$$

$$\boldsymbol{w}_k(i) = a_{kk}\boldsymbol{\psi}_k(i) + \sum_{\ell\in\mathcal{N}_k\backslash\{k\}} a_{\ell k}\overline{\boldsymbol{\psi}}_\ell(i). \quad (13)$$

B. Performance Analysis

In this subsection, we study the mean and mean-square error behavior of the EB-ATC diffusion strategy. Detailed derivations, notation definitions, as well as all the proofs are omitted due to the space constraint, and can be found in [33]. To begin with, the error vectors of each node k at time instant i are given by

$$\widetilde{\boldsymbol{\psi}}_k(i) = w^\circ - \boldsymbol{\psi}_k(i), \quad (14)$$

$$\widetilde{\boldsymbol{w}}_k(i) = w^\circ - \boldsymbol{w}_k(i). \quad (15)$$

Recall that under EB-ATC each node only combines the local updates $\{\overline{\boldsymbol{\psi}}_\ell(i)|\ell\in\mathcal{N}_k\}$ that were lastly received from its neighbors. Therefore, we also introduce the *a posterior* gap $\boldsymbol{\epsilon}_k(i)$ defined as:

$$\boldsymbol{\epsilon}_k(i) = \boldsymbol{\psi}_k(i) - \overline{\boldsymbol{\psi}}_k(i), \quad (16)$$

to capture the discrepancy between the local intermediate estimate $\boldsymbol{\psi}_k(i)$ and the estimate $\overline{\boldsymbol{\psi}}_k(i)$ that is available at neighboring nodes. We have

$$\boldsymbol{\epsilon}_k(i) = \begin{cases} 0, & \text{if } \left\|\boldsymbol{\epsilon}_k^-(i)\right\|^2_{Y_k} > \delta_k(i), \\ \boldsymbol{\epsilon}_k^-(i), & \text{otherwise}, \end{cases} \quad (17)$$

where $\boldsymbol{\epsilon}_k^-(i)$ was defined in (8). From (17), we have the following result.

Lemma 1. *The* a posterior *gap $\boldsymbol{\epsilon}_k(i)$ is bounded, and* $\|\boldsymbol{\epsilon}_k(i)\| \le \left(\frac{\delta_k}{\lambda_{\min}(Y_k)}\right)^{\frac{1}{2}}$.

Collecting the iterates $\widetilde{\boldsymbol{\psi}}_{k,i}$, $\widetilde{\boldsymbol{w}}_{k,i}$, and $\boldsymbol{\epsilon}_k(i)$ across all nodes we have,

$$\widetilde{\boldsymbol{\psi}}(i) = \operatorname{col}\left\{\left(\widetilde{\boldsymbol{\psi}}_k(i)\right)_{k=1}^N\right\}, \quad (18)$$

$$\widetilde{\boldsymbol{w}}(i) = \operatorname{col}\left\{\left(\widetilde{\boldsymbol{w}}_k(i)\right)_{k=1}^N\right\}, \quad (19)$$

$$\boldsymbol{\epsilon}(i) = \operatorname{col}\left\{\left(\boldsymbol{\epsilon}_k(i)\right)_{k=1}^N\right\}. \quad (20)$$

By substituting (16) into (13), the error vector in (15) can be expressed as

$$\widetilde{\boldsymbol{w}}_k(i) = \sum_{\ell\in\mathcal{N}_k} a_{\ell k}\widetilde{\boldsymbol{\psi}}_\ell(i) + \sum_{\ell\in\mathcal{N}_k\backslash\{k\}} a_{\ell k}\boldsymbol{\epsilon}_\ell(i), \quad (21)$$

which yields the following recursion:

$$\widetilde{\boldsymbol{w}}(i) = \boldsymbol{\mathcal{B}}(i)\widetilde{\boldsymbol{w}}(i-1) - \mathcal{A}^\mathsf{T}\mathcal{M}\boldsymbol{s}(i) + \mathcal{C}^\mathsf{T}\boldsymbol{\epsilon}(i), \quad (22)$$

where

$$\mathcal{A} = A\otimes I_M,\ \mathcal{C} = C\otimes I_M, \quad (23)$$

$$\boldsymbol{\mathcal{B}}(i) = \mathcal{A}^\mathsf{T}\left(I_{MN} - \mathcal{M}\boldsymbol{\mathcal{R}}_u(i)\right), \quad (24)$$

$$\boldsymbol{\mathcal{R}}_u(i) = \operatorname{diag}\left\{\left(\boldsymbol{u}_k(i)\boldsymbol{u}_k^\mathsf{T}(i)\right)_{k=1}^N\right\}, \quad (25)$$

$$\mathcal{M} = \operatorname{diag}\left\{\left(\mu_k I_M\right)_{k=1}^N\right\}, \quad (26)$$

$$\boldsymbol{s}(i) = \mathcal{A}^\mathsf{T}\operatorname{col}\left\{\left(\boldsymbol{u}_k(i)\boldsymbol{v}_k(i)\right)_{k=1}^N\right\}. \quad (27)$$

Theorem 1. (Mean Error Stability) *Suppose that Assumptions 1-2 hold. Then, the network mean error vector of EB-ATC, i.e., $\mathbb{E}[\widetilde{\boldsymbol{w}}(i)]$, is bounded input bounded output (BIBO) stable in steady state if the step-size μ_k is chosen such that*

$$\mu_k < \frac{2}{\lambda_{\max}(R_{u,k})}. \quad (28)$$

In addition, the block maximum norm of the network mean error is upper-bounded by

$$\frac{\alpha}{1-\beta}\cdot\max_{1\le k\le N}\left(\frac{\delta_k}{\lambda_{\min}(Y_k)}\right)^{\frac{1}{2}}, \quad (29)$$

where

$$\alpha = \max_{1\le k\le N}(1-a_{kk}),\ \beta = \|I_{MN} - \mathcal{M}\mathcal{R}_u\|_{b,\infty}. \quad (30)$$

Due to the triggering mechanism and the resulting *a posterior* gap, (20) correlates with the error vectors (18) and (19). Therefore explicitly characterizing the exact network MSD of EB-ATC is technically difficult. Instead, we study the upper bound of the network MSD. Letting $\mathcal{B} = \mathbb{E}\boldsymbol{\mathcal{B}}(i)$, and $\mu_{\max} = \max_{1\le k\le N}\mu_k$, we have the following results.

Theorem 2. (Mean-square Error Behavior) *Suppose that Assumptions 1-2 hold. Then, as $i\to\infty$, the network MSD of EB-ATC, i.e., $\mathbb{E}\|\widetilde{\boldsymbol{w}}(i)\|^2/N$, has a finite constant upper bound if the step sizes $\{\mu_k\}$ are chosen such that $\rho(\mathcal{D}) < 1$ is satisfied, where $\mathcal{D} = 2\mathbb{E}[\boldsymbol{\mathcal{B}}(i)^\mathsf{T}\otimes\boldsymbol{\mathcal{B}}(i)^\mathsf{T}]$. In addition, if $\{\mu_k\}$ are sufficiently small and also satisfy*

$$\frac{1-\frac{\sqrt{2}}{2}}{\lambda_{\min}(R_{u,k})} < \mu_k < \frac{1+\frac{\sqrt{2}}{2}}{\lambda_{\max}(R_{u,k})}, \quad (31)$$

then the matrix $\mathcal{D}$ can be approximated by $\mathcal{D}\approx\mathcal{F}$, where $\mathcal{F} = 2\mathcal{B}^\mathsf{T}\otimes\mathcal{B}^\mathsf{T} = \mathcal{D} + O(\mu_{\max}^2)$, and an upper bound of the network MSD in steady state is given by

$$\left[(f_1+f_2)^\mathsf{T}\left(I_{M^2N^2} - \mathcal{F}\right)^{-1} + f_{3,\infty}\right]\sigma + O(\mu_{\max}^2), \quad (32)$$

where,

$$f_1 = \operatorname{vec}\left(\mathcal{A}^\mathsf{T}\mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}\right),\ f_2 = 2\Delta\operatorname{vec}\left(\mathcal{C}^\mathsf{T}\mathcal{C}\right),$$

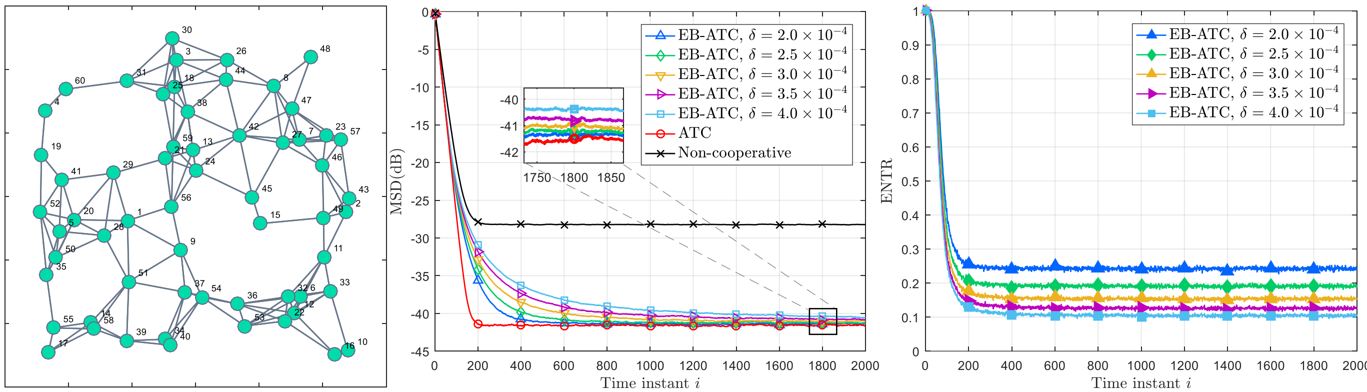
$$f_{3,\infty} = 2\lim_{i\to\infty}\sum_{j=0}^{i-1}\operatorname{vec}\left(\mathcal{C}^\mathsf{T}\mathcal{G}(i-j)\mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}\right)^\mathsf{T}\mathcal{F}^j.$$

$$\mathcal{S} = \operatorname{diag}\left\{\left(\sigma_{v,k}^2 R_{u,k}\right)_{k=1}^N\right\},$$

$$\Delta = \sum_{k=1}^N\left(\frac{\delta_k}{\lambda_{\min}(Y_k)}\right)^{\frac{1}{2}},$$

$$\mathcal{G}(i) = \mathbb{E}\operatorname{diag}\left\{\left(\boldsymbol{\gamma}_k(i)I_M\right)_{k=1}^N\right\} - I_{MN}$$

$$\sigma = \frac{1}{N}\operatorname{vec}(I_{MN}) \quad (33)$$



(a) Network topology (b) MSD performance (c) Average ENTR

Fig. 1: Simulation results for the network.

IV. Simulation Results

In this section, numerical examples are provided to illustrate the MSD performance and energy-efficiency of the proposed EB-ATC, and to compare against ATC and the non-cooperative LMS algorithm. We performed simulations on a network with $N = 60$ nodes as depicted in Fig. 1(a). The measurement noise powers $\{\sigma_{v,k}^2\}$ are generated from a uniform distribution over $[-25, -10]$ dB. We consider a parameter of interest w° with dimension $M = 10$, and suppose that the zero-mean regressor $u_k(i)$ has covariance $R_{u,k} = \sigma_{u,k}^2 I_M$, where the coefficients $\{\sigma_{u,k}^2\}$ are drawn uniformly from the interval $[1, 2]$. For the ease of implementation, we adopt constant and uniform triggering thresholds $\delta_k(i) = \delta$ for all $\{k, i\}$, and identity weighting matrix $Y_k = I_M$ for the event triggering function of every node. Moreover, we use the *Metropolis* rule [8] for the diffusion combination (13). All the simulation results are averaged over 200 Monte Carlo runs.

From Fig. 1(b), it can be observed that compared with the ATC strategy, MSDs of the proposed EB-ATC in steady state are higher by a few dBs, but still much lower than that of the non-cooperative LMS algorithm, which demonstrates the capability of EB-ATC to preserve the benefits of diffusion cooperation. We also observe that, the convergence of EB-ATC is relatively slower. This is because in the transient phase, the event-based communication mechanism of EB-ATC restricts the frequency of exchanging the newest local intermediate estimates $\{\psi_{k,i}\}$, for the purpose of energy saving.

Indeed, EB-ATC achieves significant communication overhead savings compared to ATC. To visualize this, we define the expected network triggering rate (ENTR) as follows:

$$\mathrm{ENTR}(i) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}\gamma_k(i). \qquad (34)$$

The ENTR at time instant i captures how frequently communication is triggered by each node at that time instant i, on average. ENTR is directly proportional to the average communication overhead incurred by the nodes in the network at each time instant. From (34), it is clear that $0 \le \mathrm{ENTR}(i) \le 1$, so a smaller value of $\mathrm{ENTR}(i)$ implies a lower energy consumption. Note that ATC has $\mathrm{ENTR}(i) = 1$ for all time instants i. From Fig. 1(c), we observe that the ENTR for EB-ATC decays rapidly over time during the transient phase, and for all the different triggering thresholds we tested, EB-ATC uses less than 30% of the communication overhead of ATC after the time instant $i \approx 150$, which is the average time that the MSD of ATC is within 95% of its steady-state value. This demonstrates that even though EB-ATC has not reached steady state (which occurs on average at $i \approx 600$), communication between nodes do not trigger very frequently as the intermediate estimates do not change significantly after this time instant. Furthermore, in steady state, although each node maintains estimates that are close to the true parameter value, communication triggering does not completely stop. This is due to occasional abrupt changes in the random noise and regressors, which can make the local intermediate estimate deviate significantly from its previous iteration value.

V. Conclusion

We have proposed an event-based diffusion strategy in which communication from each node to its neighbors is triggered only when a significant change has occurred in its local intermediate estimate. The proposed strategy is not only able to significantly reduce communication overhead, but can still maintain good MSD performance at steady state compared with the conventional diffusion ATC strategy. Future research includes analyzing the expected triggering rate theoretically, characterizing the rate of convergence, and establishing their relationships with the triggering threshold, so that the thresholds can be selected to optimally balance between the estimation performance and the communication energy consumption.

Acknowledgment

This work was conducted within the Delta-NTU Corporate Lab for Cyber-Physical Systems with funding support from Delta Electronics Inc. and the National Research Foundation (NRF) Singapore under the Corp Lab@University Scheme.

REFERENCES

[1] M. G. Rabbat and R. D. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 798–808, April 2005.

[2] N. Bogdanović, J. Plata-Chaves, and K. Berberidis, "Distributed incremental-based LMS for node-specific adaptive parameter estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5382–5397, Oct 2014.

[3] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, April 2010.

[4] K. Srivastava and A. Nedic, "Distributed asynchronous constrained stochastic optimization," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 772–790, Aug 2011.

[5] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, March 2010.

[6] S. Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec 2012.

[7] A. H. Sayed, S. Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.

[8] A. H. Sayed, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing*. Elsevier, 2014, vol. 3, pp. 323 – 453.

[9] ——, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, April 2014.

[10] W. Hu and W. P. Tay, "Multi-hop diffusion LMS for energy-constrained distributed estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4022–4036, Aug 2015.

[11] Y. Zhang, C. Wang, L. Zhao, and J. A. Chambers, "A spatial diffusion strategy for tap-length estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4487–4501, Sept 2015.

[12] R. Abdolee and B. Champagne, "Diffusion LMS strategies in sensor networks with noisy input data," *IEEE/ACM Trans. Netw.*, vol. 24, no. 1, pp. 3–14, Feb 2016.

[13] S. Ghazanfari-Rad and F. Labeau, "Formulation and analysis of LMS adaptive networks for distributed estimation in the presence of transmission errors," *IEEE Internet Things J.*, vol. 3, no. 2, pp. 146–160, April 2016.

[14] K. Ntemos, J. Plata-Chaves, N. Kolokotronis, N. Kalouptsidis, and M. Moonen, "Secure information sharing in adversarial adaptive diffusion networks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. PP, no. 99, pp. 1–1, 2017.

[15] C. Wang, Y. Zhang, B. Ying, and A. H. Sayed, "Coordinate-descent diffusion learning by networked agents," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 352–367, Jan 2018.

[16] J. Plata-Chaves, N. Bogdanović, and K. Berberidis, "Distributed diffusion-based LMS for node-specific adaptive parameter estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3448–3460, July 2015.

[17] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Multitask diffusion adaptation over asynchronous networks," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2835–2850, June 2016.

[18] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks with common latent representations," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 563–579, April 2017.

[19] Y. Wang, W. P. Tay, and W. Hu, "A multitask diffusion strategy with optimized inter-cluster cooperation," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 504–517, April 2017.

[20] J. Fernandez-Bes, J. Arenas-García, M. T. M. Silva, and L. A. Azpicueta-Ruiz, "Adaptive diffusion schemes for heterogeneous networks," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5661–5674, Nov 2017.

[21] X. Zhao and A. H. Sayed, "Single-link diffusion strategies over adaptive networks," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, March 2012, pp. 3749–3752.

[22] R. Arablouei, S. Werner, K. Doğançay, and Y.-F. Huang, "Analysis of a reduced-communication diffusion LMS algorithm," *Signal Processing*, vol. 117, pp. 355–361, 2015.

[23] W. Huang, X. Yang, and G. Shen, "Communication-reducing diffusion LMS algorithm over multitask networks," *Information Sciences*, vol. 382, pp. 115–134, 2017.

[24] R. Arablouei, S. Werner, Y. F. Huang, and K. Doğançay, "Distributed least mean-square estimation with partial diffusion," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 472–484, Jan 2014.

[25] M. O. Sayin and S. S. Kozat, "Compressive diffusion strategies over distributed networks for reduced communication load," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5308–5323, Oct 2014.

[26] I. E. K. Harrane, R. Flamary, and C. Richard, "Doubly compressed diffusion LMS over adaptive networks," in *Proc. 50-th Asilomar Conf. on Signals, Sys. and Comp.*, Nov 2016, pp. 987–991.

[27] J. Wu, Q. S. Jia, K. H. Johansson, and L. Shi, "Event-based sensor data scheduling: Trade-off between communication rate and estimation quality," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 1041–1046, April 2013.

[28] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Trans. Autom. Control*, vol. 60, no. 10, pp. 2661–2675, Oct 2015.

[29] A. Mohammadi and K. N. Plataniotis, "Event-based estimation with information-based triggering and adaptive update," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4924–4939, Sept 2017.

[30] G. S. Seyboth, D. V. Dimarogonas, and K. H. Johansson, "Event-based broadcasting for multi-agent average consensus," *Automatica*, vol. 49, no. 1, pp. 245–252, 2013.

[31] W. Hu, L. Liu, and G. Feng, "Consensus of linear multi-agent systems by distributed event-triggered strategy," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 148–157, Jan 2016.

[32] I. Utlu, O. F. Kilic, and S. S. Kozat, "Resource-aware event triggered distributed estimation over adaptive networks," *Digital Signal Processing*, vol. 68, pp. 127–137, 2017.

[33] Y. Wang, W. P. Tay, and W. Hu, "An event-based diffusion LMS strategy," *Available at arXiv:1803.00368*, Feb 2018.