24th Mediterranean Conference on Control and Automation (MED)
June 21-24, 2016, Athens, Greece

Distributed Drift Estimation for Time Synchronization in Lossy Networks

Miloš S. Stanković, Srdjan S. Stanković and Karl Henrik Johansson

***Abstract*— Two distributed asynchronous drift estimation algorithms for time synchronization in networks with random communication delays and measurement noise are proposed. The algorithms are superior to similar existing methods. Convergence of the algorithms to consensus in the mean square sense and w.p.1 is proved. Rate of convergence of the algorithms is analyzed. It is also shown that the algorithms can be applied as flooding algorithms, with one reference node. Illustrative simulation results are also given.**

I. Introduction

Networked Cyber-Physical Systems (CPS), Internet of Things (IoT) and Sensor Networks (SN) have emerged as research areas of paramount importance with many conceptual and practical challenges, and with numerous applications [1]–[3]. One of the basic requirements in these networked systems is *time synchronization*, *i.e.*, it is necessary for all the nodes to share a common notion of time. The problem of time synchronization has attracted a lot of attention, but still represents a challenge due to multi-hop communications, stochastic delays, communication and measurement noise, *e.g.*, [4]. An important class of time synchronization algorithms is based on full distribution of functions without reference nodes and when all the nodes run the same algorithm [5], [6]. Many distributed schemes using the gradient have been proposed (*e.g.*, [7]), including consensus-based algorithms treated in a unified way in [8]. In spite of successful solutions, efficient algorithms guaranteeing convergence in lossy networks with random delays and measurement noise are still lacking.

In this paper we propose two new *distributed drift estimation algorithms* for time synchronization in networks with random communication delays and clock reading noise. The algorithms are in the form of recursions of *distributed asynchronous stochastic approximation* type. The proposed algorithms are structurally different and simpler than analogous schemes proposed in the literature [8]–[10]. It is proved under general conditions, related to the *gossip communication protocol* and the network topology, that the proposed algorithms converge to *consensus* in the mean square sense and with probability one (w.p.1). An additional contribution of the paper is an estimation of the *convergence rate* of the algorithms which is important for their utilization in conjunction with appropriate algorithms for offset estimation. It is also demonstrated that the algorithm can be easily transformed into a *flooding* algorithm [11], with one predefined reference node. Some illustrative simulation results are also presented.

M. S. Stanković is with the Innovation Center, School of Electrical Engineering, University of Belgrade, Serbia (e-mail: milos.stankovic@ic.etf.rs).

S. S. Stanković is with the School of Electrical Engineering, University of Belgrade, Serbia and the Vlatacom Institute, Belgrade, Serbia (e-mail: stankovic@etf.rs).

K. H. Johansson is with the ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: kallej@kth.se).

This work was supported by the EU Marie Curie CIG, Knut and Alice Wallenberg Foundation and the Swedish Research Council.

II. Algorithm

Assume a communication network consisting of n nodes, formally represented by a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{E}$ the set of arcs defining the structure of inter-node communications. Assume also that $\mathcal{N}_i^+$ is the out-neighborhood and $\mathcal{N}_i^-$ the in-neighborhood of the node i.

Assume that each node contains a *local clock* whose output, defining *local time*, is given for any *absolute time* $t \in \mathcal{R}$ by

$$\tau_i(t) = \alpha_i t + \beta_i + \xi_i(t), \qquad (1)$$

where α_i is local *drift*, β_i local *offset* and $\xi_i(t)$ *measurement noise*, appearing due to equipment instabilities, round off errors, thermal noise, etc. (see, *e.g.,* [8], [12]). Each node i applies an *affine transformation* to $\tau_i(t)$, producing the *corrected local time*

$$\bar{\tau}_i(t) = a_i \tau_i(t) + b_i = g_i t + f_i + a_i \xi_i(t), \qquad (2)$$

where a_i and b_i are local *correction parameters*, $g_i = a_i \alpha_i$ is the *corrected drift* and $f_i = a_i \beta_i + b_i$ the *corrected offset*, $i = 1, \ldots, n$.

This paper deals with distributed real-time estimation of parameters a_i and b_i, aimed at asymptotically providing a *common clock*, *i.e.*, equal corrected drifts g_i and equal corrected offsets f_i, $i = 1, \ldots, n$. We assume that inter-node communication is organized according to the broadcast gossip scheme, without requiring any supervisory or fusion center. In this sense, we assume that each node $j \in \mathcal{N}$ has its own *special local broadcast clock* that ticks at the instants $t_l^j \in \mathcal{R}$, $l = 1, 2, \ldots$, according to a Poisson process with the rate μ_j, independently of the other nodes. At each tick of its broadcast clock, node j broadcasts its current local time to its out-neighbors $i \in \mathcal{N}_j^+$. In general, each node $i \in \mathcal{N}_j^+$ hears the broadcast with probability $p_{ji} > 0$; however, for the sake of clearer presentation, we assume throughout this paper that $p_{ji} = 1$. The message of node j is received by node i at the time instant

$$t_l^{j,i} = t_l^j + \delta_l^{j,i}, \qquad (3)$$

where $\delta_l^{j,i}$ represents the corresponding *communication delay* (see, *e.g.*, [13] for the presentation of physical and technical sources of the delays). All the nodes which have received the broadcast calculate their own current corrected times $\bar{\tau}_i(t_l^{j,i})$ and *update the values of their parameters* a_i and b_i. The process is repeated after each tick of a broadcast clock of any node in the network (we assume that only one clock can tick at a given time).

In this paper we pay attention only to the algorithm for updating values of the parameters a_i in (2) (due to the lack of space). In accordance with the presented gossip mechanism, we attach to the pair of nodes (j,i) the following *average error function*

$$\bar{\varphi}_i^a(t_l^j) = E\{\Delta\bar{\tau}_j(t_l^j) - \Delta\bar{\tau}_i(t_l^{j,i})\}, \tag{4}$$

where $\Delta\bar{\tau}_j(t_l^j)$ and $\Delta\bar{\tau}_i(t_l^{j,i})$ are the *increments of the corrected local times* given by $\Delta\bar{\tau}_j(t_l^j) = \bar{\tau}_j(t_l^j) - \bar{\tau}_j(t_m^j) = a_j\Delta\tau_j(t_l^j)$ and $\Delta\bar{\tau}_i(t_l^{j,i}) = a_i\Delta\tau_i(t_l^{j,i})$, where $\Delta\tau_j(t_l^j) = \tau_j(t_l^j) - \tau_j(t_m^j) = \alpha_j\Delta t_l^j + \Delta\xi_j(t_l^j)$, $\Delta t_l^j = t_l^j - t_m^j$, $\Delta\xi_j(t_l^j) = \xi_j(t_l^j) - \xi_j(t_m^j)$, $\Delta\tau_i(t_l^{j,i}) = \alpha_i\Delta t_l^{j,i} + \Delta\xi_i(t_l^{j,i})$, with $\Delta t_l^{j,i} = \Delta t_l^j + \Delta\delta_l^{j,i}$, where $\Delta\delta_l^{j,i} = \delta_l^{j,i} - \delta_m^{j,i}$, and $\Delta\xi_i(t_l^{j,i}) = \xi_i(t_l^{j,i}) - \xi_i(t_m^{j,i})$; $m \in \{1,\ldots,l-1\}$ is an index to be specified later. We assume that the communication delay can be decomposed as

$$\delta_l^{j,i} = \bar{\delta}^{j,i} + \eta_i(t_l^{j,i}),$$

where $\bar{\delta}^{j,i}$ is assumed to be constant, while $\eta_i(t_l^{j,i})$ represents a stochastically time varying component, so that $\Delta\delta_l^{j,i} = \Delta\eta_i(t_l^{j,i})$, where $\Delta\eta_i(t_l^{j,i}) = \eta_i(t_l^{j,i}) - \eta_i(t_m^{j,i})$.

We propose the following updating rule for node i, based on current realizations of (4):

$$\hat{a}_i(t_l^{j,i+}) = \hat{a}_i(t_l^j) + \varepsilon_i(t_l^j)\gamma_{ij}\varphi_i^a(t_l^j), \tag{5}$$

where γ_{ij} are *a priori* adopted nonnegative weights expressing relative importance of the communication between the nodes j and i (see the discussion below, Subsection III.D), $\varphi_i^a(t_l^j) = \Delta\hat{\tau}_j(t_l^j) - \Delta\hat{\tau}_i(t_l^{j,i})$, $\Delta\hat{\tau}_j(t_l^j) = \Delta\bar{\tau}_j(t_l^j)|_{a_j=\hat{a}_j(t_l^j)}$ and $\Delta\hat{\tau}_i(t_l^{j,i}) = \Delta\bar{\tau}_i(t_l^{j,i})|_{a_i=\hat{a}_i(t_l^j)}$, $\hat{a}_i(t_l^j)$ represents the old estimate, $\hat{a}_i(t_l^{j,i+})$ the new estimate, while $\varepsilon_i(t_l^j)$ is a positive step size. The updating procedure (5) generates, in such a way, a recursion of distributed asynchronous stochastic approximation type, where (4) plays the role of the local regression function [14]. It will be assumed that the initial estimates are defined as $\hat{a}_i(t_0) = 1$ where t_0 is a time instance before the first time any node i receives a message and updates its parameter estimate.

The updating rule (5) can be rewritten in terms of the corrected drift $\hat{g}_i(\cdot) = \hat{a}_i(\cdot)\alpha_i$ and noise terms:

$$\hat{g}_i(t_l^{j,i+}) = \hat{g}_i(t_l^j) + \varepsilon_i(t_l^j)\gamma_{ij}\psi_i^a(t_l^j), \tag{6}$$

where

$$\psi_i^a(t_l^j) = \alpha_i\{[\hat{g}_j(t_l^j) - \hat{g}_i(t_l^j)]\Delta t_l^j + \frac{1}{\alpha_j}\hat{g}_j(t_l^j)\Delta\xi_j(t_l^j) - \frac{1}{\alpha_i}\hat{g}_i(t_l^j)\Delta\xi_i(t_l^{j,i}) - \hat{g}_i(t_l^j)\Delta\eta_i(t_l^{j,i})\}.$$

Remark 1: The basic drift estimation scheme represented by (5) and (6) is autonomous, independent from offset estimation. Its role is analogous to the role of the distributed drift estimation schemes described in [8], [9]. However, it does not belong to the class of the so-called CBTS algorithms discussed in [8]: it is structurally different and simpler (CBTS algorithms contain explicit *relative drift* estimation, with additional dynamics and nonlinearities). ■

In order to obtain a global model for the whole network, we introduce a *global virtual clock* with the rate equal to $\mu = \sum_{i=1}^n \mu_i$, that ticks whenever any of the local broadcast clocks ticks. Let the k-th tick of the virtual clock, $k = 1, 2, \ldots$, correspond to the k-th update of the parameter estimates (k-th iteration). Let j be the index of the node that broadcasts at the k-th tick. We replace the variable t_l^j by the index k in all the above defined functions of time, *e.g.*, $\tau_j(t_l^j) = \tau_j(k)$, $\xi_j(t_l^j) = \xi_j(k)$, etc. We also replace $t_l^{j,i}$ by k, so that $\tau_i(t_l^{j,i}) = \tau_i(k)$, $\xi_i(t_l^{j,i}) = \xi_i(k)$, etc.

The vector of all the corrected (modified) drifts in the network $\hat{g}(k) = [\hat{g}_1(k)\cdots\hat{g}_n(k)]^T$, where $\hat{g}_i(k) = \hat{a}_i(k)\alpha_i$ ($\hat{a}_i(k) = \hat{a}_i(t_l^j)$), satisfies the following model

$$\hat{g}(k+1) = \hat{g}(k) + \varepsilon(k)Z(k)\hat{g}(k), \tag{7}$$

where $\hat{g}(k+1) = [\hat{g}_1(t_l^{j,1+})\cdots\hat{g}_n(t_l^{j,n+})]^T$, $\varepsilon(k) = \mathrm{diag}\{\varepsilon_1(k),\ldots,\varepsilon_n(k)\}$, $\varepsilon_i(k) = \varepsilon_i(t_l^j)$ (see (5)),

$$Z(k) = A\Gamma(k)\Delta t(k) + N_g(k),$$

$A = \mathrm{diag}\{\alpha_1,\ldots,\alpha_n\}$, $\Gamma(k) = [\Gamma(k)_{\mu\nu}]$, with $\Gamma(k)_{\mu\mu} = -\gamma_{\mu j}$ and $\Gamma(k)_{\mu j} = \gamma_{\mu j}$ for all $\mu \in \mathcal{N}_j^+$, with $\Gamma(k)_{\mu\nu} = 0$ otherwise, $\Delta t(k) = t_l^j - t_m^j$, while the noise term is defined

$$N_g(k) = -A\Gamma_d(k)\Delta\eta_d(k) + A\Gamma(k)\Delta\xi_d(k)A^{-1},$$

where $\Delta\eta_d(k) = \mathrm{diag}\,\Delta\eta(k)$, $\Delta\eta(k) = [\Delta\eta_1(k)\cdots\Delta\eta_n(k)]^T$, $\Delta\xi_d(k) = \mathrm{diag}\,\Delta\xi(k)$, $\Delta\xi(k) = [\Delta\xi_1(k)\cdots\Delta\xi_n(k)]^T$, $\Gamma_d(k) = \mathrm{diag}\{\gamma_{1j},\cdots,\gamma_{nj}\}$ ($\Delta\eta_\mu(k) = 0$, $\Delta\xi_\mu(k) = 0$, $\gamma_{\mu j} = 0$ for $\mu \notin \mathcal{N}_j^+$).

In the sequel, we shall make distinction between *two versions* of the proposed drift estimation algorithm, depending on the choice of $\Delta t(k)$:

ALGa: $m = l-1$ in (4), so that $\Delta t(k) = t_l^j - t_{l-1}^j$ with $E\{\Delta t(k)\} = \frac{1}{\mu_j}$ and $E\{\Delta t(k)^2\} = \frac{2}{\mu_j^2}$;

ALGb: $m = 1$ in (4), so that $\Delta t(k) = t_l^j - t_1^j$ with $E\{\Delta t(k)\} = \frac{1}{\mu}k$, since $E\{\Delta t(k)\} = \frac{n_j}{\mu_j} = \frac{k}{\mu}$, where n_j is the expected number of broadcasts of node j in the interval $t_l^j - t_1^j$; also, $E\{\Delta t(k)^2\} = 2k\frac{1}{\mu^2}$.

Remark 2: Two versions of the drift estimation algorithm are proposed in order to indicate two important ways of applying the adopted design methodology. In this sense, ALGa has its roots in the ATS algorithm proposed in [9], and ALGb in the LSTS algorithm proposed in [8], [10]. A pseudo periodic version of ALGa has been presented for the first time in [15]. ■

III. Convergence Analysis

A. Preliminaries

Within the exposed general setting, we additionally assume:

(A1) Graph $\mathcal{G}$ has a spanning tree.

(A2) $\{\xi_i(k)\}$ and $\{\eta_i(k)\}$, $i = 1, \ldots n$, are mutually independent zero mean i.i.d. random sequences, bounded w.p.1.

(A3) The step size $\varepsilon_i(k)$ is defined as $\varepsilon_i(k) = \nu_i(k)^{-\zeta}$, where $\nu_i(k) = \sum_{m=1}^{k} I\{i \in \mathcal{N}_j^+\}$ represents the number of updates of node i up to the instant k, $I\{\cdot\}$ denotes the indicator function, while $\zeta = \zeta'$ for ALGa, $\zeta = 1 + \zeta'$ for ALGb, where $\frac{1}{2} < \zeta' \leq 1$.

Remark 3: (A1) and (A2) are standard assumptions [16], [17]. (A3) is very important for practice, since it eliminates the need for a centralized clock which defines the common step size as a function of k. The choice of the step size parameter ζ is different for ALGa and ALGb, as a consequence of the different properties of the random variable $\Delta t(k)$. In ALGa, when $\Delta t(k)$ is m.s. bounded, we have the "classical" choice of ζ [14], while in ALGb, when $\Delta t(k)$ diverges to infinity, we have a somewhat unusual step size which prevents divergence of the corresponding regression function, and contributes, at the same time, to noise immunity by introducing decreasing weights to the noise terms (see the analysis below). ■

We first clarify asymptotics of the step size.

Lemma 1: [18] Let (A1) and (A3) be satisfied, let p_i be the unconditional probability of node i *to update* its parameters, $i = 1, \ldots, n$ and let $\zeta > 0$; then, for a given $q \in (0, \frac{1}{2})$, there exists such a $\bar{k}$ that w.p.1 for all $k \geq \bar{k}$

$$\varepsilon(k) = \frac{1}{k^\zeta} P^{-\zeta} + \tilde{\varepsilon}(k), \tag{8}$$

where $P = \mathrm{diag}\{p_1, \ldots, p_n\}$, $\tilde{\varepsilon}(k) = \mathrm{diag}\{\tilde{\varepsilon}_1(k), \ldots, \tilde{\varepsilon}_n(k)\}$, $|\tilde{\varepsilon}_i(k)| \leq \tilde{\varepsilon}_i \frac{1}{k^{\zeta + \frac{1}{2} - q}}$, $0 < \tilde{\varepsilon}_i < \infty$, $i = 1, \ldots, n$. ■

Properties of matrix $\bar{\Gamma} = E\{\Gamma(k)\}$ are essential for convergence: it contains information about the network structure and the weights of particular links. It has the structure of the weighted Laplacian matrix for $\mathcal{G}$:

$$\bar{\Gamma} = \begin{bmatrix} -\sum\limits_{j, j \neq 1} \gamma_{1j}\pi_{1j} & \gamma_{12}\pi_{12} & \cdots & \gamma_{1n}\pi_{1n} \\ \gamma_{21}\pi_{21} & -\sum\limits_{j, j \neq 2} \gamma_{2j}\pi_{2j} & \cdots & \gamma_{2n}\pi_{2n} \\ & & \ddots & \\ \gamma_{n1}\pi_{n1} & \gamma_{n2}\pi_{n2} & \cdots & -\sum\limits_{j, j \neq n} \gamma_{nj}\pi_{nj} \end{bmatrix}$$

($\gamma_{ij} = 0$ when $j \notin \mathcal{N}_i^-$), where π_{ij} is the probability that the node i updates its parameters as a consequence of a tick of node j (π_{ij} follows, in general, from the Poisson rate μ_j and the transmission probability p_{ij}; as stated above, in this paper we have $\pi_{ij} = \pi_j$, $i \in \mathcal{N}_j^+$, where π_j is the probability of node j to broadcast).

B. Convergence to Consensus

Theorem 1: Let assumptions (A1), (A2) and (A3) be satisfied. Then $\hat{g}(k)$ generated by (7) converges to $\hat{g}_\infty = \chi^* \mathbf{1}$ in the mean square sense and w.p.1, where χ^* is a random variable with bounded second moment.

Proof: Let $B(k) = P^{-\zeta} A \Gamma(k)$ and $\bar{B} = E\{B(k)\} = P^{-\zeta} A \bar{\Gamma}$. Having in mind that $\bar{\Gamma}$ has one eigenvalue at the origin and the remaining ones in the left half plane [17], we introduce $T = \left[\mathbf{1} \vdots T_{n \times (n-1)} \right]$, where $T_{n \times (n-1)}$ is an $n \times (n-1)$ matrix, such that $\mathrm{span}\{T_{n \times (n-1)}\} = \mathrm{span}\{\bar{B}\}$. Consequently, $T^{-1} = \begin{bmatrix} v \\ \hdashline T'_{(n-1) \times n} \end{bmatrix}$, where v represents the left eigenvector of $\bar{B}$ corresponding to the zero eigenvalue, satisfying $\|v\| = 1$, while $T'_{(n-1) \times n}$ follows from the definition of T. Then, $T^{-1} \bar{B} T = \begin{bmatrix} 0 & 0_{1 \times (n-1)} \\ 0_{(n-1) \times 1} & \bar{B}^* \end{bmatrix}$, where $\bar{B}^*$ is Hurwitz; also,

$$T^{-1} B(k) T = \begin{bmatrix} 0 & B_1(k)^* \\ 0_{(n-1) \times 1} & B_2(k)^* \end{bmatrix}, \tag{9}$$

where $B_2(k)^*$ is a $(n-1) \times (n-1)$ matrix.

After coming back to (7), we first insert $\varepsilon(k)$ given in (8). Then, we introduce $\tilde{g}(k) = T^{-1} \hat{g}(k)$ and decompose $\tilde{g}(k)$ as $\tilde{g}(k) = [\tilde{g}(k)^{[1]} \vdots \tilde{g}(k)^{[2]T}]^T$, where $\tilde{g}(k)^{[1]} = \tilde{g}_1(k)$ and $\tilde{g}(k)^{[2]} = [\tilde{g}_2(k) \cdots \tilde{g}_n(k)]^T$. After neglecting the higher order terms from (8), we obtain

$$\tilde{g}(k+1)^{[1]} = \tilde{g}(k)^{[1]} + \frac{1}{k^\zeta} F_1(k) \Delta t(k) \tilde{g}(k)^{[2]} + \frac{1}{k^\zeta} H_1(k) \tilde{g}(k) \tag{10}$$

$$\tilde{g}(k+1)^{[2]} = \{I + \frac{1}{k^\zeta} [\bar{B}^* + F_2(k)] \Delta t(k)\} \tilde{g}(k)^{[2]} + \frac{1}{k^\zeta} H_2(k) \tilde{g}(k), \tag{11}$$

where $F_1(k)$ and $F_2(k)$ are $1 \times (n-1)$ and $(n-1) \times (n-1)$ matrices, respectively, defined by

$$T^{-1}[B(k) - \bar{B}] T = \begin{bmatrix} 0 & F_1(k) \\ 0_{(n-1) \times 1} & F_2(k) \end{bmatrix},$$

while $H_1(k)$ and $H_2(k)$ are $1 \times n$ and $(n-1) \times n$ matrices, respectively, defined by

$$T^{-1} P^{-\zeta} N_g(k) T = \begin{bmatrix} H_1(k) \\ \hdashline H_2(k) \end{bmatrix}.$$

Introduce two Lyapunov functions $V^g(k) = E\{\|\tilde{g}(k)^{[1]}\|^2\}$ and $W^g(k) = E\{\tilde{g}(k)^{[2]T} R^g \tilde{g}(k)^{[2]}\}$, where R^g is a symmetric positive definite matrix satisfying

$$R^g \bar{B}^* + \bar{B}^{*T} R^g = -Q^g, \tag{12}$$

for any given $Q^g > 0$, having in mind that $\bar{B}^*$ is Hurwitz [17].

In order to obtain an estimate of $V^g(k)$, we iterate (10) back to the initial condition, and obtain that

$$\tilde{g}(k+1)^{[1]} = \tilde{g}_1(k+1)^{[1]} + \tilde{g}_2(k+1)^{[1]}, \tag{13}$$

where

$$\tilde{g}_1(k+1)^{[1]} = \Pi(k,1)^{[1]} \tilde{g}(1)^{[1]} \tag{14}$$

and

$$\tilde{g}_2(k+1)^{[1]} = \sum_{\sigma=1}^{k} \frac{1}{\sigma^{\zeta}} \Pi(k,\sigma+1)^{[1]} [F_1(\sigma)\Delta t(\sigma) + H_1(\sigma)^{[2]}] \tilde{g}(\sigma)^{[2]}, \quad (15)$$

while $\Pi(k,l)^{[1]} = \prod_{\sigma=l}^{k}(1+\frac{1}{\sigma^{\zeta}}H_1(\sigma)^{[1]})$, $\Pi(k,k+1)^{[1]} = 1$; $H_1(k)^{[1]}$ is a scalar following from the decomposition $H_1(k) = [H_1(k)^{[1]} \vdots H_1(k)^{[2]}]$. Therefore,

$$V^g(k+1) \leq 2V_1^g(k+1) + 2V_2^g(k+1), \quad (16)$$

where $V_1^g(k+1) = E\{\tilde{g}_1(k+1)^{[1]2}\}$ and $V_2^g(k+1) = E\{\tilde{g}_2(k+1)^{[1]2}\}$.

Introduce subsequences $\{\kappa^j(v)\}$, $j = 1,\ldots,n$, $v = 1,2,\ldots$ of the set of positive integers $\mathcal{I}^+$ ($\cup_j\{\kappa^j(v)\} = \mathcal{I}^+$), in which each $\kappa^j(v)$ defines an instant k corresponding to a broadcast of node j ($\kappa^j(v_1) < \kappa^j(v_2)$ for $v_1 < v_2$). Moreover, each of these subsequences can be decomposed into two sub-subsequences $\{\kappa_o^j(v)\}$ and $\{\kappa_e^j(v)\}$, corresponding to odd and even indices v; define also ordered subsequences $\{\kappa_o(v)\} = \cup_j\{\kappa_o^j(v)\}$ and $\{\kappa_e(v)\} = \cup_j\{\kappa_e^j(v)\}$. Accordingly, we define $\Pi(k,1)_o^{[1]} = \prod_{\sigma\in\{\kappa_o(v)\}}(1+\frac{1}{\sigma^{\zeta}}H_1(\sigma)^{[1]})$ and $\Pi(k,1)_e^{[1]} = \prod_{\sigma\in\{\kappa_e(v)\}}(1+\frac{1}{\sigma^{\zeta}}H_1(\sigma)^{[1]})$.

In the case of ALGa, the sequences $\{H_1(\sigma)\}$ for both $\sigma \in \{\kappa_o(v)\}$ and $\sigma \in \{\kappa_e(v)\}$ are composed of zero mean i.i.d. random variables (see the properties of $N_g(k)$ in (7)). Therefore, it is straightforward to show (using *e.g.*, [19]) that $\sup_k E\{(\Pi(k,1)_o^{[1]})^2\} < \infty$ and $\sup_k E\{(\Pi(k,1)_e^{[1]})^2\} < \infty$.

In the case of ALGb, according to the corresponding form of $N_g(k)$, we have $H_1(\sigma) = \tilde{H}_1(\sigma) - \tilde{H}_1(\sigma_1^j)$ for $\sigma \in \{\kappa^j(v)\}$, where $\{\tilde{H}_1(\sigma)\}$ is a zero mean i.i.d. random sequence and $\tilde{H}_1(\sigma_1^j)$ a random variable satisfying (A2) (notice that $\sigma_1^j = \kappa^j(1)$). Obviously, for $\sigma \in \{\kappa^j(v)\}$

$$E\{(1-\frac{1}{\sigma^{1+\zeta'}}H_1(\sigma))^2|\mathcal{F}_{\sigma_1^j}\} \leq 1 - c_1\frac{1}{\sigma^{1+\zeta'}} + c_2\frac{1}{\sigma^{2(1+\zeta')}},$$

where $\mathcal{F}_{\sigma_1^j}$ is a minimal σ-algebra generated by $\tilde{H}_1(\sigma_1^j)$ and $0 \leq c_1, c_2 < \infty$. Having in mind (A1), we conclude that, for $\sigma \in \{\kappa^j(v)\}$ and every $j \in \{1,\ldots,n\}$, $\sup_\sigma E\{(\Pi(\sigma,1)^{[1]})^2\} < \infty$.

Therefore, it follows that $\sup_k V_1^g(k) < \infty$ for both ALGa and ALGb.

We analyze $V_2^g(k)$ reasoning in a similar way.

For ALGa, we write $\tilde{g}_2(k+1)^{[1]} = \tilde{g}_2^o(k+1)^{[1]} + \tilde{g}_2^e(k+1)^{[1]}$, where $\tilde{g}_2^o(k+1)^{[1]}$ contains the terms from the sum at the right hand side of (15) corresponding to $\sigma \in \{\kappa_o(v)\}$, and $\tilde{g}_2^e(k+1)^{[1]}$ the terms corresponding to $\sigma \in \{\kappa_e(v)\}$. Both $\tilde{g}_2^o(k+1)^{[1]}$ and $\tilde{g}_2^e(k+1)^{[1]}$ represent weighted sums of random variables $F_1(\sigma)^{[2]}\Delta t(\sigma) + H_1(\sigma)^{[2]}$, for $\sigma \in \{\kappa_o(v)\}$ and $\sigma \in \{\kappa_e(v)\}$, respectively, that are i.i.d., due to the nature of $N_g(k)$ (sequence $\{F_1(k)\}$ is zero mean i.i.d. by basic assumptions of the gossip broadcast). Therefore, we obtain, after some technicalities, that

$$V_2^g(k+1) \leq C_1 \sum_{\sigma=1}^{k} \frac{1}{\sigma^{1+q''}} W^g(\sigma), \quad (17)$$

where $C_1 > 0$ and $q'' > 0$ and are constants.

The same conclusion holds for ALGb, after taking into account (A3) and the boundedness condition from (A2). Namely, the sum at the right hand side of (15) contains now the term $H_1(\sigma) = \tilde{H}_1(\sigma) - \tilde{H}_1(\sigma_1^j)$. The sequence $\{F_1(\sigma)\Delta t(\sigma) + \tilde{H}_1(\sigma)^{[2]}\}$ is zero mean i.i.d., producing, within the calculation of $V_2^g(k)$, a term satisfying an inequality analogous to (17) (with an appropriate constant). The remaining term $\tilde{H}_1(\sigma_1^j)$ is bounded according to (A2), and produces a term satisfying also an inequality analogous to (17) (having in mind that $\sum_\sigma \frac{1}{\sigma^{(1+\zeta')}} < \infty$).

Consequently, for both ALGa and ALGb,

$$V^g(k+1) \leq C_2[1 + \max_{1\leq\sigma\leq k} W_g(\sigma)], \quad (18)$$

where $C_2 > 0$, having in mind that $\sum_{k=1}^{\infty}\frac{1}{\sigma^{1+q''}} < \infty$ in (17).

Estimation of $W^g(k)$ will start from iterating back (11) to the initial condition and the following decomposition $\tilde{g}(k+1)^{[2]} = \tilde{g}_1(k+1)^{[2]} + \tilde{g}_2(k+1)^{[2]}$, where

$$\tilde{g}_1(k+1)^{[2]} = \Pi(k,1)^{[2]}\tilde{g}(1)^{[2]} \quad (19)$$

and

$$\tilde{g}_2(k+1)^{[1]} = \sum_{\sigma=1}^{k} \frac{1}{\sigma^{\zeta}} \Pi(k,\sigma+1)^{[2]} H_2(\sigma)^{[2]} \tilde{g}(\sigma)^{[2]}, \quad (20)$$

where $\Pi(k,l)^{[2]} = \prod_{\sigma=l}^{k}[I + \frac{1}{\sigma^{\zeta}}(\bar{B}^* + F_2(\sigma))]$, with $\Pi(k,k+1)^{[2]} = I$. Therefore,

$$W^g(k+1) \leq 2W_1^g(k+1) + 2W_2^g(k+1)$$

where $W_1^g(k+1) = E\{\tilde{g}_1(k+1)^{[2]T}R^g\tilde{g}_1(k+1)^{[2]}\}$ and $W_2^g(k+1) = E\{\tilde{g}_2(k+1)^{[2]T}R^g\tilde{g}_2(k+1)^{[2]}\}$.

In the case of ALGa, we start the analysis by observing that for any n-vector x and any $\sigma \in [1,k]$

$$x^T E\{\Pi(\sigma,\sigma)^{[2]T}R^g\Pi(\sigma,\sigma)^{[2]}\}x \leq [1 - \frac{1}{\sigma^{\zeta'}} \times \frac{\lambda_{min}(Q^g)}{\max_j \mu_j \lambda_{max}(R^g)} + \frac{1}{\sigma^{2\zeta'}}\frac{\lambda_{max}(R^g)}{\min_j \mu_j \lambda_{min}(R^g)}\bar{\lambda}]x^T R^g x, \quad (21)$$

where $0 < \lambda_{min}(Q^g), \lambda_{min}(R^g), \lambda_{max}(R^g), \bar{\lambda} < \infty$, $\bar{\lambda} = \lambda_{max}(E\{[\bar{B}^* + F_2(\sigma)^{[2]}]^T[\bar{B}^* + F_2(\sigma)^{[2]}]\})$. Therefore, we obtain, having in mind that $\{F_2(k)\}$ is a zero mean i.i.d. sequence, that for k large enough

$$W_1^g(k+1) \leq \prod_{\sigma=1}^{k}(1 - c_3\frac{1}{\sigma^{\zeta'}})C_3, \quad (22)$$

where $0 < c_3, C_3 < \infty$; consequently, $W_1^g(k) \to_{k\to\infty} 0$.

The term $W_2^g(k)$ can be analyzed in the case of ALGa similarly as $V_2^g(k)$. We write $\tilde{g}_2(k+1)^{[2]} = \tilde{g}_2^o(k+1)^{[2]} + \tilde{g}_2^e(k+1)^{[2]}$, where $\tilde{g}_2^o(k+1)^{[2]}$ contains the terms from the sum at the RHS of (20) corresponding to $\sigma \in \{\kappa_o(v)\}$, and $\tilde{g}_2^e(k+1)^{[2]}$ the terms corresponding to $\sigma \in \{\kappa_e(v)\}$. Since sequences $\{H_2(\sigma)^{[2]}\}$ are in both cases zero mean i.i.d. with finite second moments and $H_2(\sigma)^{[2]}$ is independent of $\tilde{g}_2(\sigma)^{[2]}$, we obtain that $W_2^g(k+1) \leq 2W_{2,o}^g(k+1) + 2W_{2,e}^g(k+1)$, where

$$W_{2,o}^g(k+1) \leq C_4 \sum_{\sigma\in\{\kappa_o(v)\},\sigma\leq k} \frac{1}{\sigma^{2\zeta'}} \prod_{s\in\{\kappa_o(v)\},\sigma<s\leq k}(1-c_4\frac{1}{s^{\zeta'}})W_2^g(\sigma) \quad (23)$$

where $0 < c_4, C_4 < \infty$ ($W_{2,e}^g(k+1)$ satisfies (23) with appropriate constants, after replacing $\{\kappa_o(v)\}$ by $\{\kappa_e(v)\}$). It follows from (23), having in mind [19] and the classical results from stochastic approximation [14], that

$$\max_{1\leq\sigma\leq k+1} W_2^g(\sigma) \leq \alpha^W(k) \max_{1\leq\sigma\leq k} W_2^g(\sigma), \quad (24)$$

where $\lim_{k\to\infty}\alpha^W(k) = 0$. Consequently, $\lim_{k\to\infty} W_2^g(k) = 0$, so that, we conclude that $\sup_k W^g(k) < \infty$ and $\lim_{k\to\infty} W^g(k) = 0$.

The same result concerning $W^g(k)$ holds also for ALGb. Namely, the inequalities related to $W_1^g(k)$ remain the same as in the case of ALGa, as (21) also holds in the case of ALGb. In the case of $W_2^g(k)$, it is clear that the analysis related to $V_2^g(k)$ in the context of ALGb can again be applied. Namely, we introduce $H_2(\sigma) = \tilde{H}_2(\sigma) - \tilde{H}_2(\sigma_1^j)$, $\sigma \in \{\kappa^j(v)\}$, $j = 1,\ldots,n$, and observe that $\{\tilde{H}_2(\sigma)\}$ is a zero mean i.i.d. sequence. The term in $W_2^g(k)$ corresponding to $\tilde{H}_2(\sigma)$ satisfies an inequality like (23). In relation with the term containing $\tilde{H}_2(\sigma_1^j)$, we apply (A2) and observe that $\lim_{v\to\infty}\sum_{\nu=1}^{v}\frac{1}{\kappa^j(\nu)^{(1+\zeta')}}\prod_{u=\nu+1}^{v}(1-c_6\frac{1}{\kappa^j(u)^{\zeta'}}) = 0$, $j = 1,\ldots,n$, $c_6 > 0$. Thus, for ALGb we also have $\lim_{k\to\infty} W^g(k) = 0$.

Using the result related to $W^g(k)$, we come back to the inequality (18), and obtain that $\sup_k V^g(k) < \infty$. According to [19] and the arguments exposed therein, we conclude that $\tilde{g}(k)^{[1]}$ tends to a random variable χ^* ($E\{\chi^{*2}\} < \infty$) and $\tilde{g}(k)^{[2]}$ to zero in the mean square sense and w.p.1. Consequently

$$\hat{g}_\infty = T\begin{bmatrix}\lim_{k\to\infty}\tilde{g}(k)^{[1]}\\ \hdashline 0\end{bmatrix} = \chi^*\mathbf{1}, \quad (25)$$

which proves the theorem. ■

C. Convergence Rate

Rate of convergence to consensus of drift estimation schemes is of great importance for distributed time synchronization [8]. Estimation of the rate of convergence to consensus will be based on $\tilde{g}(k)^{[2]}$ in (11), using the methodology of [14].

Theorem 2: Let (A1), (A2) and (A3) hold. Then, $z(k) = k^{\zeta d}\tilde{g}(k)^{[2]}$, where $d > 0$ and $\tilde{g}(k)^{[2]}$ is defined by (11), converges to zero in the mean square sense and w.p.1 for: a) ALGa: $\zeta d = \zeta' d < \zeta' - \frac{1}{2}$ when $\zeta' < 1$, and $d < \min(\frac{1}{2}, \frac{\lambda_{min}(Q^g)}{\max_j \mu_j \lambda_{max}(R^g)})$ when $\zeta' = 1$; b) ALGb: $\zeta d < \zeta'$ when $\zeta' < 1$, and $d < \min(\frac{1}{2}, \frac{1}{2}\frac{\lambda_{min}(Q^g)}{\mu\lambda_{max}(R^g)})$ when $\zeta' = 1$.

Proof: After introducing the expression for $z(k)$ into (11), we obtain (see [14])

$$z(k+1) = (1+\frac{1}{k})^{\zeta d}\{z(k) + \frac{1}{k^\zeta}[\bar{B}^* + F_2(k)] \times \Delta t(k) z(k)\} + \frac{1}{k^{\zeta(1-d)}}H_2(k)\tilde{g}(k). \quad (26)$$

We use here the approximation $(1+\frac{1}{k})^{\zeta d} \approx 1 + \zeta d\frac{1}{k}$ and obtain, after neglecting higher order terms, that

$$z(k+1) = z(k) + \{\frac{1}{k^\zeta}[\bar{B}^* + F_2(k)]\Delta t(k) + \zeta d\frac{1}{k}I\}z(k) + \frac{1}{k^{\zeta(1-d)}}H_2(k)\tilde{g}(k) \quad (27)$$

Now we can apply directly the methodology of the proof of Theorem 1 to the analysis of (27). In the case of ALGa, when $\zeta = \zeta'$, for $\zeta' < 1$ we obtain directly that $z(k)$ converges to zero in the mean square sense and w.p.1 provided $\sum_{k=1}^{\infty}\frac{1}{k^{2\zeta'(1-d)}} < \infty$. When $\zeta' = 1$, we have a change in the basic contraction condition from Theorem 1, which becomes $\frac{\lambda_{min}(Q^g)}{\max_j \mu_j \lambda_{max}(R^g)} > \zeta' d$; together with $d < \frac{1}{2}$, this proves the part related to ALGa. In the case of ALGb, for $\zeta' < 1$ we have convergence of $z(k)$ to zero provided $(1+\zeta')(1-d) > 1$ (having in mind the definition of $H_2(k)$, *i.e.*, the existence of a bounded term depending on the noise realizations at some initial time instants - see the related discussion in the proof of Theorem 1). For $\zeta' = 1$, we have the following inequalities: $\frac{\lambda_{min}(Q^g)}{\mu\lambda_{max}(R^g)} > 2d$ and $d < \frac{1}{2}$, having in mind that $\zeta = 2$. ■

Remark 4: The value ζd is higher for ALGb for a given common ζ', indicating superiority of ALGb in practice and its advantage for real-time applications within offset estimation schemes [8], [15]. ■

D. Weight Selection: Flooding

The choice of weights γ_{ij} in the error function (4) and of the Poisson rates μ_j is dictated, in general, by the relative importance of particular communication lines and/or nodes in the network. If one wishes to express high confidence in the precision of a given clock j, there are two main implementation ways: 1) to increase either the elements of the j-th column γ_{ij}, $i = 1,\ldots,n$, or the Poisson rate μ_j ; 2) to decrease the weights γ_{ji}, $i = 1,\ldots,n$. The second way leads to lowering increments of the local parameter changes at node j, and, therefore, to lower influence of the rest of the network to the corresponding local parameter estimates. In the limit, this leads to the situation in which node j does not update its parameters ($\gamma_{ij} = 0$, $i = 1,\ldots,n$), and becomes a *reference node* with fixed characteristics; the whole algorithm becomes, in this sense, an algorithm of *flooding* type (see [11], [20]).

Corollary 1: Let the assumptions of Theorem 1 be satisfied. Let the node λ be a center node in $\mathcal{G}$, with the corrected drift $\hat{g}_\lambda^*$. Then, after setting $\mathcal{N}_\lambda^- = \emptyset$ (or $\gamma_{\lambda i} = 0$, $i = 1,\ldots,n$), algorithm (5) provides convergence of all corrected drifts $\hat{g}_i(k)$, $i = 1,\ldots,n$, $i \neq \lambda$, to $\hat{g}_\lambda$ in the mean square sense and w.p.1.

Proof: After reducing $\mathcal{N}_\lambda^-$ to an empty set, the underlying graph still satisfies (A1), since the new graph still has a spanning tree with the node λ as the center node. The rest of the proof follows directly Theorem 1. As the recursion at node λ reduces to $\hat{a}_\lambda(k+1) = \hat{a}_\lambda(k)$, it follows that $\lim_{k\to\infty}\hat{g}_\lambda(k) = \hat{g}_\lambda(0) = \hat{g}_\lambda^*$, and as all the nodes have the same limit, this limit is equal to $\hat{g}_\lambda^*$. ■

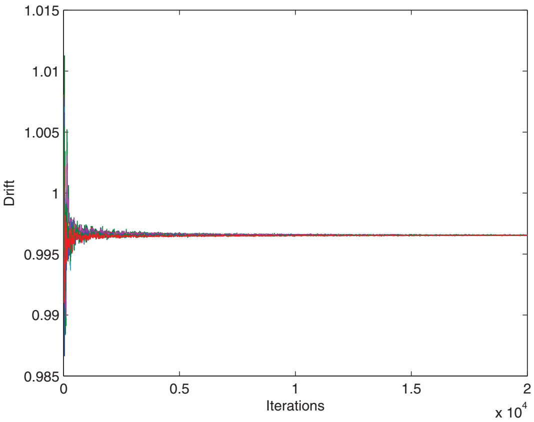

Fig. 1. Drift estimates: ALGa

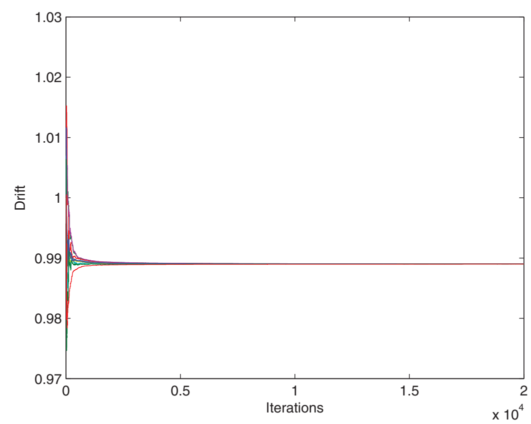

Fig. 2. Drift estimates: ALGb

IV. SIMULATIONS

Simulation experiments have been undertaken in order to clarify different aspects of the proposed distributed algorithm. A network of ten nodes forming a directed graph has been simulated, with α_i and β_i randomly chosen in the intervals $(0.999, 1.001)$ and $(-0.001, 0.001)$, respectively. Average communication delays $\bar{\delta}^{j,i}$ have been chosen at random in the interval $(0.005, 1.005)$, while $\{\eta(k)\}$ and $\{\xi(k)\}$ have been simulated as Gaussian white noise zero mean sequences with standard deviation equal to 0.001. It has been adopted that $\zeta' = 0.99$.

Typical behavior of the drift estimates generated by ALGa and ALGb in the presence of stochastic delays and measurement noise is illustrated in Figs. 1 and 2, respectively. Convergence to consensus can be clearly observed; relative advantage of ALGb is evident.

V. CONCLUSION

In this paper two algorithms of asynchronous stochastic approximation type are proposed for drift estimation in distributed time synchronization schemes in lossy networks with random communication delays and clock measurement noise. Starting from general assumptions related to the properties of the network and the communication protocol based on broadcast gossip, it has been proved that the proposed algorithms ensure asymptotic consensus of the drift estimates in the mean square sense and w.p.1. The rate of convergence to consensus is estimated for both algorithms. Convergence is also proved for an algorithm of flooding type obtained by taking the corrected drift of one node as a reference.

An immediate challenge is to investigate a distributed *offset estimation algorithm* based on the drift estimates generated by the proposed algorithms.

REFERENCES

[1] K.-D. Kim and P. R. Kumar, “Cyber–physical systems: A perspective at the centennial,” *Proc. IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, 2012.

[2] J. Holler, V. Tsiatsis, C. Mulligan, S. Avesand, S. Karnouskos, and D. Boyle, *From Machine-to-machine to the Internet of Things: Introduction to a New Age of Intelligence*. Academic Press, 2014.

[3] I. F. Akyildiz and M. C. Vuran, *Wireless sensor networks*. John Wiley & Sons, 2010.

[4] B. Sundaraman, U. Buyand, and A. Kshemkalyani, “Clock synchronization for wireless sensor networks: a survey,” *Ad Hoc Networks*, vol. 3, pp. 281–323, 2005.

[5] O. Simeone, U. Spagnolini, Y. Bar-Ness, and S. H. Strogatz, “Distributed synchronization in wireless networks,” *IEEE Signal Process. Mag.*, vol. 25, pp. 81–97, 2008.

[6] R. Solis, V. Borkar, and P. R. Kumar, “A new distributed time synchronization protocol for multi hop wireless networks,” in *Proc. IEEE Conf. Decision and Control*, 2006, pp. 2734–2739.

[7] P. Sommer and R. Wattenhofer, “Gradient clock synchronization in wireless sensor networks,” in *Proc. Int. Conf. Inf. Proc. in Sensor Networks*, 2009, pp. 37–48.

[8] Y.-P. Tian, S. Zong, and Q. Cao, “Structural modeling and convergence analysis of consensus-based time synchronization algorithms over networks: non-topological conditions,” *Automatica*, vol. 65, pp. 64–75, 2016.

[9] L. Schenato and F. Fiorentin, “Average timesynch: a consensus-based protocol for time synchronization in wireless sensor networks,” *Automatica*, vol. 47, no. 9, pp. 1878–1886, 2011.

[10] Y.-P. Tian, “LSTS: a new time synchronization protocol for networks with random communication delays,” in *Proc. IEEE Conf. Decision and Control*, 2015, pp. 7404–7409.

[11] M. Maroti, B. Kusy, G. Simon, and A. Ledeczi, “The flooding time synchronization protocol,” in *Proc. 2nd ACM Intern. Conf. Embedded Netw. Contr. Syst., SenSys’04*, 2004, pp. 39–49.

[12] C. Liao and P. Barooah, “Distributed clock skew and offset estimation from relative measurements in mobile networks with Markovian switching topologies,” *Automatica*, vol. 49, pp. 3015–3022, 2013.

[13] G. Xiong and S. Kishore, “Analysis of distributed consensus time synchronization with Gaussian delay over wireless sensor networks,” *EURASIP J. on Wireless Comm. and Netw.*, no. 1, pp. 1–9, 2009.

[14] H. F. Chen, *Stochastic approximation and its applications*. Dordrecht, the Netherlands: Kluwer Academic, 2002.

[15] M. S. Stanković, S. S. Stanković, and K. H. Johansson, “Distributed time synchronization in lossy wireless sensor networks,” in *Proc. 3rd IFAC Workshop on Distr. Estim. Contr. Netw. Syst., NECSYS12*, vol. 3, 2012, pp. 25–30.

[16] R. Olfati-Saber, A. Fax, and R. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proc. IEEE*, vol. 95, pp. 215–233, 2007.

[17] M. S. Stanković, S. S. Stanković, and K. H. Johansson, “Distributed blind calibration in lossy sensor networks via output synchronization,” *IEEE Trans. Autom. Control*, vol. 60, pp. 3257–3262, 2015.

[18] A. Nedic, “Asynchronous broadcast-based convex optimization over a network,” *IEEE Trans. Autom. Control*, vol. 56, pp. 1337–1351, 2011.

[19] M. Huang and J. H. Manton, “Stochastic consensus seeking with noisy and directed inter-agent communications: fixed and randomly varying topologies,” *IEEE Trans. Autom. Control*, vol. 55, pp. 235–241, 2010.

[20] K. S. Yildirim, R. Carli, and L. Schenato, “Adaptive control-based clock synchronization in wireless sensor networks,” in *Proc. European Contr. Conf. ECC15*, 2015.