Clock Synchronization and Estimation in Highly Dynamic Networks: An Information Theoretic Approach

Ofer Feinerman[1,⋆] and Amos Korman[2]

[1] The Shlomo and Michla Tomarin Career Development Chair, The Weizmann Institute of Science, Rehovot, Israel
ofer.feinerman@weizmann.ac.il

[2] CNRS and University Paris Diderot, Paris, 75013, France
amos.korman@liafa.univ-paris-diderot.fr

Abstract. We consider the *External Clock Synchronization* problem in dynamic sensor networks. Initially, sensors obtain inaccurate estimations of an external time reference and subsequently collaborate in order to synchronize their internal clocks with the external time. For simplicity, we adopt the *drift-free* assumption, where internal clocks are assumed to tick at the same pace. Hence, the problem is reduced to an estimation problem, in which the sensors need to estimate the initial external time. In this context of distributed estimation, this work is further relevant to the problem of collective approximation of environmental values by biological groups.

Unlike most works on clock synchronization that assume static networks, this paper focuses on an extreme case of highly dynamic networks. We do however impose a restriction on the dynamicity of the network. Specifically, we assume a non-adaptive scheduler adversary that dictates an arbitrary, yet *independent*, meeting pattern. Such meeting patterns fit, for example, with short-time scenarios in highly dynamic settings, where each sensor interacts with only few other arbitrary sensors.

We propose an extremely simple clock synchronization (or an estimation) algorithm that is based on weighted averages, and prove that its performance on any given independent meeting pattern is highly competitive with that of the best possible algorithm, which operates without any resource or computational restrictions, and further knows the whole meeting pattern in advance. In particular, when all distributions involved are Gaussian, the performances of our scheme coincide with the optimal performances. Our proofs rely on an extensive use of the concept of Fisher information. We use the Cramér-Rao bound and our definition of a *Fisher Channel Capacity* to quantify information flows and to obtain lower bounds on collective performance. This opens the door for further rigorous quantifications of information flows within collaborative sensors.

⋆ O.F. has been supported in part by the Clore Foundation, the Israel Science Foundation (FIRST grant no. 1694/10) and the Minerva Foundation. A.K. has been supported in part by the ANR project DISPLEXITY. This work has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No 648032).

C. Scheideler (Ed.): SIROCCO 2015, LNCS 9439, pp. 16–30, 2015.
DOI: 10.1007/978-3-319-25258-2_2

1 Introduction

1.1 Background and Motivation

Representing and communicating information is a main interest of theoretical distributed computing. However, such studies often seem disjoint from what may be the largest body of work regarding coding and communication: Information theory [7,33]. Perhaps the main reason for this stems from the fact that distributed computing studies are traditionally concerned with noiseless models of communication, in which the content of a message that passes from one node to another is not distorted. This reliability in transmission relies on an implicit assumption that error-corrections is guaranteed by a lower level protocol that is responsible for implementing communication. Indeed, when bandwidth is sufficiently large, one can encode a message with a large number of error-correcting bits in a way that makes communication noise practically a non-issue.

In some distributed scenarios, however, distortion in communication is unavoidable. One example concerns the classical problem of *clock synchronization*, which has attracted much attention from both theoreticians in distributed computing [2,25,22,30], as well as engineers [10,15,34], see [32,37,24,39] for comprehensive surveys. In this problem, processors need to synchronize their clocks (either among themselves only or with respect to a global time reference) relying on relative time measurements between clocks. Due to unavoidable unknown delays in communication, such measurements are inherently noisy. Furthermore, since the source of the noise is the delays, error-correction does not seem to be of any use for reducing the noise. The situation becomes even more complex when processors are mobile, preventing them from reducing errors by averaging repeated measurements to the same processors, and from contacting reliable processors. Indeed, the clock synchronization problem is particularly challenging in the context of wireless sensor networks and ad hoc networks which are typically formed by autonomous, and often mobile, sensors without central control.

Distributed computing models which include noisy communication call for a rigorous comprehensive study that employs information theoretical tools. Indeed, a recent trend in the engineering community is to view the clock synchronization problem from a signal processing point of view, and adopt tools from information theory (e.g., the Cramér-Rao bound) to bound the affect/impact of inherent noise [6,15], see [39] for a survey. However, this perspective has hardly received any attention by theoreticians in distributed computing that mostly focused on worst case message delays [2,25,22,4], which do not seem to be suitable for information theoretic considerations. In fact, very few works on clock synchronisation consider a system with random delays and analyse it following a rigorous theoretical distributed algorithmic type of analysis. An exception to that is the work of Lenzen el al. [23], but also that work does not involve information theory. In this current paper, we study the clock synchronization problem through the purely theoretical distributed algorithmic perspective while adopting the signal processing and information theoretic point of view. In particular, we adopt tools from Fisher Information theory [35,40].

We consider the *external* version of the problem [8,28,30,37] in which processors (referred to as sensors hereafter) collaborate in order to synchronize their clocks with an external *global clock*. Informally, sensors initially obtain inaccurate estimates of a global (external) time $\tau^* \in \mathcal{R}$ reference, and subsequently collaborate to align their internal clocks to be as close as possible to the external clock. To this end, sensors communicate through uni-directional pairwise interactions that include inherently *noisy measurements* of the relative deviation between their internal clocks and, possibly, some complementary information. To focus on the problems caused by the initial inaccurate estimations of τ^* and the noise in the communication we restrict our attention to *drift-free* settings [2,25], in which all clocks tick at the same rate. This setting essentially reduces the problem to the problem of estimating τ^*. See, e.g., [14,36,38] for works on estimation in the engineering community. In this context of distributed estimation, our model is further relevant to collective approximation of environmental values by biological groups [19,26].

With very few exceptions that effectively deal with dynamic settings [9,20], almost all works on clock synchronization (and distributed estimation) considered static networks. Indeed, the construction of efficient clock synchronization algorithms for dynamic networks is considered as a very important and challenging task[1] [32,37]. This paper addresses this challenge by considering highly dynamic networks in which sensors have little or no control on who they interact with. Specifically, we assume a non-adaptive scheduler adversary that dictates in advance a meeting-pattern for the sensors. However, the adversary we assume is not unlimited. Specifically, in this initial work[2] we restrict the adversary to provide *independent-meeting patterns* only, in which it is guaranteed that whenever a sensor views another sensor, their transitive histories are disjoint[3]. Although they are not very good representatives of communication in static networks, independent meeting patterns fit well with highly stochastic communication patterns during short-time scales, in which each sensor observes only few other arbitrary sensors (see discussion in Section 2). Given such a meeting-pattern, we are concerned with minimizing the deviation of each internal clock from the global time.

As our objective is to model small and simple sensors, we are interested in algorithms that employ elementary computations and economic use of communication. We use competitive analysis to evaluate the performances of algorithms, comparing them to the best possible algorithm that operates under the most

[1] For example, dynamic meeting patterns prevent the use of classical external clock synchronization algorithms (e.g., [27,30]) that are based on one or few *source* sensors that obtain accurate estimation of the global time and govern the synchronization of other sensors.

[2] We assume independence for simplicity. As evident by this work, the independent case is already rather complex. We leave it to future work to handle more complex dependent scenarios.

[3] Another informal way to view such patterns is that they guarantee that, given the global time, whenever a sensor views another sensor, their local clocks are independent; see Section 2 for a formal definition.

liberal version of the model that allows for unrestricted resources in terms of memory and communication capacities, and individual computational ability.

Due to space considerations, throughout this paper, most proofs are omitted. These proofs can be found in [12].

1.2 Our Contribution

Lower Bounds on Optimal Performance. We first consider algorithm **Opt**, the best possible algorithm operating on the given independent meeting pattern. We note that specifying **Opt** seems challenging, especially since we do not assume a prior distribution on the starting global time, and hence the use of Bayesian statistics seems difficult. Fortunately, for our purposes, we are merely interested in lower bounding the performances of that algorithm. We achieved that by relating the smallest possible variance of a sensor at a given time to the largest possible *Fisher Information (FI)* of the sensor at that time. This measure quantifies the sensor's current knowledge regarding the relative deviation between its local time and the global time. We provide a recursive formula to calculate J_a, the FI at sensor a, for any sensor a. Specifically, initially, the FI at a sensor is the FI in the distribution family governing its initial deviation from the global time (see Section 2 for the formal definitions). When sensor a observes sensor b, the FI at a after this observation (denoted by J'_a) satisfies:

$$J'_a \leq J_a + \frac{1}{\frac{1}{J_b} + \frac{1}{J_N}}, \tag{1}$$

where J_N is the Fisher Information in the noise distribution related to the observation. To obtain this formula we prove a generalized version of the *Fisher information inequality* [35,40]. Relying on the *Cramér-Rao bound* [7], this formula is then used to bound the corresponding variance under algorithm **Opt**. Specifically, the variance of the internal clock of sensor a is at least $1/J_a$.

Equation 1 provides immediate bounds on the convergence time. Specifically, the inequality sets a bound of J_N for the increase in the *FI* per interaction. In analogy to Channel Capacity as defined by Shannon [7] we term this upper bound as the *Fisher Channel Capacity.* Given small $\epsilon > 0$, we define the convergence time $T(\epsilon)$ as the minimal number of observations required by the typical sensor until its variance drops below ϵ^2 (see Section 2 for the formal definition). Let J_0 denote the median initial Fisher Information of sensors. Based on the Fisher Channel Capacity we prove the following.

Theorem 1. *Let $J_0 \ll 1/\epsilon^2$ for some $\epsilon > 0$. Then $T(\epsilon) \geq (\frac{1}{\epsilon^2} - J_0)/J_N$.*

A Highly Competitive Elementary Algorithm. We propose a simple clock synchronization algorithm and prove that its performance on any given independent meeting pattern is highly competitive with that of the optimal one. That is, estimations of global time at each sensor remain unbiased throughout the execution and the variance at any given time is Δ_0-competitive with the best possible variance, where Δ_0 is initial Fisher-tightness (see definition in Section 2). In contrast to the optimal algorithm that may be based on transmitting complex functions in each interaction, and on performing complex internal computations, our simple algorithm is based on far more basic rules. First, transmission is restricted to a single *accuracy* parameter. Second, using the noisy measurement of deviation from the observed sensor, and the accuracy of that sensor, the observing sensor updates its internal clock and accuracy parameter by careful, yet elementary, weighted-averaging procedures.

Our weighted-average algorithm is designed to maximize the flow of Fisher Information in interactions. This is proved by showing that the accuracy parameter is, at all times, both representative of the reciprocal of the sensor's variance and close to the Fisher Information upper bound. In short, we prove the following.

Theorem 2. *There exists a simple weighted-average based clock synchronization algorithm which is Δ_0-competitive (at any sensor and at any time).*

We note that our algorithm does not require the use of sensor identities and can thus be also employed in *anonymous* networks [1,11], yielding the same performances.

Two important corollaries of Theorem 2 follow directly from the definition of the initial Fisher-tightness Δ_0.

Corollary 1. *If the number of distributions governing the initial clocks is a constant (independent of n), then our algorithm is $O(1)$-competitive, at any sensor and at any time.*

Corollary 2. *If all distributions involved are Gaussians, then the variances of our algorithm coincide with those of the optimal one, for each sensor and at any time.*

2 Preliminaries

We consider a collection of n sensors that collaborate in order to synchronize their internal clocks with an external global clock reference. We consider a set $\mathcal{F}$ of sufficiently smooth (see definition in Section 2), probability density distributions (*pdf*) centered at zero. One specific distribution among the *pdf*s in $\mathcal{F}$ is the *noise* distribution, referred to as $N(\eta)$. Each sensor a is associated with a distribution $\Phi_a(x) \in \mathcal{F}$ which governs the initialization deviation of its internal clock from the global time as described in the next paragraph. Depending on the specific model, we assume that sensor a knows various properties of Φ_a. In the most restricted model, sensor a knows only the variance of Φ_a and in the most liberal model (considered for the sake of lower bounds), a knows the full description of Φ_a. Execution is initiated when the global time is some $\tau^* \in \mathcal{R}$, chosen by an adversary.

Two important cases are (1) when $\mathcal{F}$ contains a constant number of distributions (independent of the number of sensors) and (2) when all distributions in F are Gaussian. Both cases serve as reasonable assumptions for realistic scenarios. For the former case we shall show asymptotically optimal performances and for the latter case we shall show strict optimal (non-asymptotical) performance.

Local Clocks. Each sensor a is initialized with a local clock $\ell_a(0) \in \mathcal{R}$, randomly chosen according to $\Phi_a(x - \tau^*)$, independently of all other sensors. That is, as $\Phi_a(x)$ is centred around zero, the initial local time $\ell_a(0)$ is distributed around τ^*, and this distribution is governed by Φ_a. We stress that sensor a does not know the value τ^* and from its own local perspective the execution started at time $\ell_a(0)$. Sensors rely on both social interactions and further environmental cues[4] to improve their estimates of the global time. In between such events sensors are free to perform "shift" operations to adjust their local clocks. To focus on the problems occurred by the initial inaccurate estimations of τ^* and the noise in the communication we restrict our attention to *drift-free* settings [2,25], in which all clocks tick at the same rate, consistent with the global time.

Opinions. The drift-free assumption reduces the external clock-synchronization problem to the problem of estimating τ^*. Indeed, recall that local clocks are initialized to different values but progress at the same rate. Because sensor a can keep the precise time since the beginning of the execution, its deviation from the global time can be corrected had it known the difference between, $\ell_a(0)$, the initial local clock of a, and τ^*, the global time when the execution started. Hence, one can view the goal of sensor a as estimating τ^*. That is, without loss of generality, we may assume that all shifts performed by sensor a throughout the execution are shifts of its initial position $\ell_a(0)$ aiming to align it to be as close as possible to τ^*. Taking this perspective, we associate with each sensor an *opinion* variable x_a, initialized to $x_a(0) := \ell_a(0)$, and the goal of a is to have its opinion be as close as possible to τ^*. We view the opinion x_a as an *estimator* of τ^*, and note that initially, due to the properties of Φ_a, this estimator is unbiased, i.e., $\text{mean}(x_a(0) - \tau^*) = 0$. It is required that at any point in the execution, the opinion x_a remains an unbiased estimator of τ^*, and the goal of a is to minimize its Mean Square Error (MSE).

Due to this simple relation between internal clocks and opinions, in the remaining of this paper, we shall adopt the latter perspective and concern ourselves only with optimizing the opinions of sensors as estimators for τ^*, without discussing further the internal clocks.

Rounds. For simplicity of presentation, we assume that the execution proceeds in discrete rounds. We stress however that the rounds represent the order in which communication events occur (as determined by the meeting-pattern, see

[4] In order for the model to include environmental cues, one or more of the sensors can be taken to represent the global clock. The initial times of these sensors are chosen according to highly concentrated distributions, Φ_a, around τ^* and remain fixed thereafter.

below), and do not necessarily correspond to the actual time. Given an algorithm A, the opinion maintained by the algorithm at round t (where t is a non-negative integer) at sensor a is denoted by $x_a(t, A)$. As mentioned, the algorithm aims to keep this value as close as possible to τ^*. When A is clear from the context, we may omit writing it and use the term $x_a(t)$ instead.

In each round $t \geq 1$, a sensor may first choose to shift (or not) its opinion, and then, if specified in the meeting pattern, it observes another specified sensor, thus obtaining some information. To summarize, in each round, a sensor executes the following consecutive actions: (1) Perform internal computation; (2) Perform an opinion-shift: $x_a(t) = x_a(t-1) + \Delta(x)$; and (3) Observe (or not) another sensor. For simplicity, all these three operations are assumed to occur instantaneously, that is, in zero time.

Mobility and Adversarial Independent Meeting Patterns. In cases where sensors are embedded in a Euclidian space, distances between positioning of sensors may impact the possible interactions. To account for physical mobility, and be as general as possible, we assume that an oblivious adversary controls the meeting pattern. That is, the adversary decides (before the execution starts), for each round, which sensor observes which other sensor.

A model that includes an unlimited adversary that controls the meeting pattern appears to be too general. In this preliminary work on the subject, we restrict the adversary to provide only *independent* meeting patterns, in which the set of sensors in the transitive history of each observing sensor is disjoint from the one of the observed sensor.

Formally, given a pattern of meetings $\mathcal{P}$, sensor a and round t, we first define the set of *relevant* sensors of a at time t, denoted by $\mathcal{R}_a(t, \mathcal{P})$. At time zero, we define $\mathcal{R}_a(0, \mathcal{P}) := \{a\}$, and at round t, $\mathcal{R}_a(t, \mathcal{P}) := \mathcal{R}_a(t-1, \mathcal{P}) \cup \mathcal{R}(b, t-1, \mathcal{P})$ if a observes b at time $t-1$ (otherwise $\mathcal{R}_a(t, \mathcal{P}) := \mathcal{R}_a(t-1, \mathcal{P})$). A meeting pattern $\mathcal{P}$ is called *independent* if whenever some sensor a observes a sensor b at some time t, then $\mathcal{R}_a(t-1, \mathcal{P}) \cap \mathcal{R}(b, t-1, \mathcal{P}) = \emptyset$. Note that an independent meeting pattern guarantees that given τ^*, the internal clocks of two interacting sensors are independent. However, given τ^* and the internal clock of a, the internal clock of b and the relative time measurement between them are dependent.

Note that independent-meeting patterns are not very good representatives of communication in static networks[5]. On the other hand, independent meeting patterns fit well with highly stochastic short-time scales communication patterns,

[5] Indeed, in such patterns a sensor will not contact the same sensor twice, which contradicts many natural communication schemes in static networks. We note, however, that in some cases, a sequence of multiple consecutive observations between sensors can be compressed into a single observation of higher accuracy thus reducing the dependencies between observations, and possibly converting a dependent meeting pattern into an independent one. For example, if sensors have unique identities and sensor a observes sensor b several times is a row, and it is guaranteed that sensor b did not change its state during these observations, then these observations can be treated by a as a single, more accurate, observation of b.

in which each sensor observes only few other arbitrary sensors. In this sense, such patterns can be considered as representing an extreme case of dynamic systems.

Because sensors have no control of when their next interaction will occur, or if it will occur at all, we require that estimates at each sensor be as accurate as possible at *any* point in time. This requirement is stronger than the liveness property that is typically required from distributed algorithms [21].

Convergence Time. Consider a meeting pattern $\mathcal{P}$. Given small $\epsilon > 0$, the *convergence time* $T(\epsilon)$ of an algorithm A is defined as the minimal number of observations made by the typical sensor until its variance is less than ϵ^2. More formally, let ρ denote the first round when we have more than half of the population satisfying $\text{var}(X_a(t, A)) < \epsilon^2$. For each sensor a, let $R(a)$ denote the number of observations made by a until time ρ. The convergence time $T(\epsilon)$ is defined as the median of $R(a)$ over all sensors a. Note that $T(\epsilon)$ is a lower bound on ρ, since since each sensor observes at most one sensor in a round.

Communication. We assume that sensors are anonymous and hence, in particular, they do not know who they observe. Conversely, for the sake of lower bounds, we allow a much more liberal setting, in which sensors have unique identifiers and know who they interact with.

When a sensor a observes another sensor b at some round t, the information transferred in this interaction contains a *passive* component and, possibly, a complementary *active* one. The passive component is a noisy relative deviation measurement between their opinions:

$$\tilde{d}_{ab}(t) = x_b(t) - x_a(t) + \eta,$$

where the additive noise term, η, is chosen from the noise probability distribution $N(\eta) \in \mathcal{F}$ whose variance is known to the sensors. (Note that this measurement is equivalent to the relative deviation measurement between the sensors' current local times because all clocks tick at the same pace.)

Elementary Algorithms. Our reference for evaluating performances is algorithm **Opt** which operates under the most liberal version of our model, which carries no restrictions on memory, communication capacities or internal computational power, and provides the best possible estimators at any sensor and at any time (we further assume that sensors acting under **Opt** know the meeting pattern in advance). In general, algorithm **Opt** may use complex calculations over very wasteful memories that include detailed distribution density functions, and possibly, accumulated measurements. Our main goal is to identify an algorithm whose performance is highly competitive with that of **Opt** but wherein communication and memory are economically used, and the local computations simple. Indeed, when it comes to applications to tiny and limited processors, simplicity and economic use of communication are crucial restrictions.

An algorithm is called *elementary* if the internal state of each sensor a contains a constant number of real[6] numbers, and the internal computations that a sensor can perform consist of a constant number of basic arithmetic operations, namely: addition, subtraction, multiplication, and division.

Competitive Analysis. Fix a finite family $\mathcal{F}$ of smooth *pdf*'s centered at zero (see the definition for smoothness in the next paragraph), and fix an assignment of a distribution $\Phi_a \in \mathcal{F}$ to each sensor a. For an algorithm A and an independent meeting pattern $\mathcal{P}$, let $X_a(t, A, \mathcal{P})$ denote the random variable indicating the opinion of sensor a at round t. Let $\text{mean}(X_a(t, A, \mathcal{P}))$ and $\text{var}(X_a(t, A, \mathcal{P}))$ denote, respectively, the mean and variance of $X_a(t, A, \mathcal{P})$, where these are taken over all possible random initial opinions, communication errors, and possibly, coins flipped by the algorithm. Note that the unbiased assumption requires that $\text{mean}(X_a(t, A, \mathcal{P})) = \tau^*$. An algorithm A is called λ-competitive, if for *any* independent pattern of meetings $\mathcal{P}$, *any* sensor a, and at *any* time t, we have: $\text{var}(X_a(t, A, \mathcal{P})) \leq \lambda \cdot \text{var}(X_a(t, \mathbf{Opt}, \mathcal{P}))$.

Fisher Information and the Cramér-Rao Bound. The Fisher information is a standard way of evaluating the amount of information that a set of random measurements holds about an unknown parameter τ of the distribution from which these measurements were taken. We provide some definitions for this notion; for more information the reader may refer to [7,40].

A single variable probability distribution function (*pdf*) Φ is called *smooth* if it satisfies the following conditions, as stated by Stam [35]: (1) $\Phi(x) > 0$ for any $x \in \mathcal{R}$, (2) the derivative Φ' exists, and (3) the integral $\int \frac{1}{\Phi(y)}(\Phi'(y))^2 dy$ exists, i.e., $\Phi'(y) \to 0$ rapidly enough for $|y| \to \infty$. Note that, in particular, these conditions hold for natural distributions such as the Gaussian distribution. Recall that we consider a finite set $\mathcal{F}$ of smooth one variable *pdf*s, one of them being the noise distribution $N(\eta)$, and all of which are centered at zero.

For a smooth *pdf* Φ, let $J^\tau_\Phi := \int \frac{1}{\Phi(y)}(\Phi'(y))^2 dy$ denote the Fisher information in the parameterized family $\{(\Phi(x, \tau)\}_{\tau\in\mathcal{R}} = \{(\Phi(x - \tau)\}_{\tau\in\mathcal{R}}$ with respect to τ. In particular, let $J_N = J^\tau_N$ denote the Fisher information in the parameterized family $\{N(\eta - \tau)\}_{\tau\in\mathcal{R}}$. More generally, consider a multi-variable *pdf* family $\{(\Phi(z_1 - \tau, z_2 \ldots z_k))\}_{\tau\in\mathcal{R}}$ where τ is a translation parameter. The Fisher information in this family with respect to τ is defined as: $J^\tau_\Phi = \int \frac{1}{\Phi(z_1-\tau,z_2\ldots z_k)} \left[\frac{d\Phi(z_1-\tau,z_2\ldots z_k)}{d\tau} \right]^2 dz_1, dz_2 \ldots dz_k$ if the integral exists. As previously noted [40], since τ is a translation parameter, Fisher information is both unique (there is no freedom in choosing the parametrization) and independent of τ.

The Fisher information derives its importance by association with the Cramér-Rao inequality [7]. This inequality lower bounds the variance of the best possible

[6] We assume real numbers for simplicity. It seems reasonable to assume that when sufficiently accurate approximation is stored instead of the real numbers similar results could be obtained.

estimator of τ^* by the reciprocal of the Fisher information that corresponds to the random variables on which this estimator is based.

Theorem 3. **[The Cramér-Rao inequality]** *Let $\hat{X}$ be any unbiased estimator of $\tau^* \in \mathcal{R}$ which is based on a multi-variable sample $\bar{z} = (z_1, z_2 \dots z_k)$ taken from $\Phi(z_1 - \tau^*, z_2 \dots z_k)$. Then* $\mathrm{var}(\hat{X}) \geq 1/J_\Phi^\tau$.

Initial Fisher-Tightness: To define the initial Fisher-tightness parameter Δ_0, we first define the *Fisher-tightness* of a single variable smooth distribution Φ centered at zero, as $\Delta(\Phi) = \mathrm{var}(\Phi) \cdot J_\Phi^\tau$. Note that, by the Cramér-Rao bound, $\Delta(\Phi) \geq 1$ for any such distribution Φ. Moreover, equality holds if Φ is Gaussian [7]. Recall that $\mathcal{F}$ is the finite collection of the smooth distributions containing the distributions Φ_a governing the initial opinions of sensors. The *initial Fisher-tightness* Δ_0 is the maximum of the Fisher-tightness over all distributions in $\mathcal{F}$ and the noise distribution. Specifically, let $\Delta_0 = \max\{\Delta(\Phi) \mid \Phi \in \mathcal{F}\}$. Two important observations are:

- If $\mathcal{F}$ contains a constant number of distributions then Δ_0 is a constant.
- If the distributions in $\mathcal{F}$ are all Gaussians then $\Delta_0 = 1$.

3 Lower Bounds on the Variance of Opt

In this section we provide lower bounds on the performances of algorithm **Opt** over a fixed independent pattern of meetings $\mathcal{P}$. Note that we are interested in bounding the performances of **Opt** and not in specifying its instructions. Identifying the details of **Opt** may still be of interest, but it is beyond the scope of this paper.

For simplicity of presentation, we assume that the rules of **Opt** are deterministic. We note, however, that our results can easily be extended to the case that **Opt** is probabilistic. For simplicity of notations, since this section deals only with algorithm **Opt** acting over $\mathcal{P}$, we use variables, such as the opinion $X_a(t)$ and the memory $Y_a(t)$ of sensor a, without parametrizing them by neither **Opt** nor by $\mathcal{P}$.

Under algorithm **Opt**, we assume that each sensor holds initially, in addition to the variance of Φ_a, the precise functional form of the distribution Φ_a (recall, Φ_a is centered at zero). In addition, we assume that sensors have unique identifiers and that each sensor knows the whole pattern $\mathcal{P}$ in advance. Moreover, we assume that each sensor a knows for each other sensor b, the *pdf* Φ_b governing b's initial opinion. All this information is stored in one designated part of the memory of a.

Since **Opt** does not have any bandwidth constrains, we may assume, without loss of generality, that whenever some sensor a observes another sensor b, it obtains the whole memory content of b. Since **Opt** is deterministic, its previous opinion-shifts can be extracted from its interaction history, which is, without loss of generality, encoded in its memory[7]. Hence, when sensor a observes sensor

[7] In case **Opt** is probabilistic, previous shifts can be extracted from the memory plus the results of coin flips which may be encoded in the memory of the sensor as well.

b at some round t, and receives b's memory together with the noisy measurement $\tilde{d}_{ab}(t) = x_b(t) - x_a(t) + \eta$, sensor a may extract all previous opinion-shifts of both itself and b, treating the measurement $\tilde{d}_{ab}(t)$ as a noisy measurement of the deviation between the initial opinions, i.e., $\tilde{d}_{ab}(0) = x_b(0) - x_a(0) + \eta$. In other words, to understand the behavior of **Opt** at round t, one may assume that sensors never shift their opinions until round t, when they use all memory they gathered to shift their opinion in the best possible manner[8]. It follows that apart from the designated memory part that all sensors share, the memory $M_a(t)$ of sensor a at round t contains the initial opinion $X_a(0)$ and a collection $Y_a(t-1) := \{\tilde{d}_{bc}(0)\}_{bc}$ of relative deviation measurements between initial opinions. That is, $M_a(t) = (X_0(t), Y_a(t-1))$. This multi-valued memory variable $M_a(t)$ contains all the information available to a at round t. In turn, this information is used by the sensor to obtain its opinion $X_a(t)$ which is required to serve as an unbiased estimator of τ^*.

The Fisher Information of Sensors. We now define the notion of the Fisher Information associated with a sensor a at round t. This definition will be used to bound from below the variance of $X_a(t)$ under algorithm **Opt**.

Consider the multi-valued memory variable $M_a(t) = (X_0(t), Y_a(t-1))$ of sensor a that at round t. Note that $Y_a(t-1)$ is independent of τ^*. Indeed, once the adversary decides on the value τ^*, all sensors' initial opinions are chosen with respect to τ^*. Hence, since sensors' memories contains only relative deviations between opinions, the memories by themselves do not contain any information regarding τ^*. In contrast, given τ^*, the random variables $Y_a(t-1)$ and $X_a(0)$ are, in general, dependent. Furthermore, in contrast to $Y_a(t-1)$, the value of $X_a(0)$ depends on τ^*, as it is chosen according to $\Phi_a(x-\tau^*)$. Hence, $M_a(t)$ is distributed according to a *pdf* family $\{(m_a(t), \tau)\}$ parameterized by a translation parameter τ. Based on $M_a(t)$, the sensor produces an unbiased estimation $X_a(t)$ of τ^*, that is, it should hold that: $\text{mean}(X_a(t) - \tau^*) = 0$, where the mean is taken with respect to the distribution of the random multi-variable $M_a(t)$.

Definition: The *Fisher Information (FI)* of sensor a at round t, termed $J_a(t)$, is the the Fisher information in the parameterized family $\{(m_a(t), \tau)\}_{\tau \in \mathcal{R}}$ with respect to τ.

By the Cramér-Rao bound, the variance of any unbiased estimator used by the sensor a at round t is bounded from below by the reciprocal of the *FI* of sensor a at that time. That is, we have:

Lemma 1. *$\text{var}(X_a(t)) \geq 1/J_a(t)$.*

[8] This observation implies, in particular, that previous opinion-shifts of sensors do not affect subsequent estimators in a way that may cause a conflict (a conflict may arise, e.g., when optimizing one sensor at one time necessarily makes estimators at another sensor, at a later time, sub-optimal), hence algorithm **Opt** is well-defined.

3.1 An Upper Bound on the Fisher Information $J_a(t)$

Lemma 1 implies that lower bounds on the variance of the opinion of a sensor can be obtained by bounding from above the corresponding *FI*. To this end, we prove the following recursive inequality. To establish the proof we had to extends the Fisher information inequality [35,40] to our multi-variable (possibly dependent) convolution case.

Theorem 4. *The FI of sensor a under algorithm* **Opt** *satisfies:* $J_a(t+1) \leq J_a(t) + 1/(\frac{1}{J_b(t)} + \frac{1}{J_N})$.

4 A Highly-Competitive Elementary Algorithm

We define an elementary algorithm, termed **ALG**, and prove that its performances are highly-competitive with those of **Opt**. In this algorithm, each sensor a stores in its memory a single parameter $c_a \in \mathcal{R}$ that represents its *accuracy* regarding the quality of its current opinion with respect to τ^*. The initial accuracy of sensor a is set to $c_a(0) = 1/\mathrm{var}(\Phi_a)$. When sensor a observes sensor b at some round t, it receives $c_b(t)$ and $\tilde{d}_{ab}(t)$, and acts as follows. Sensor a first computes the value $\hat{c}_b(t) = c_b(t)/(1 + c_b(t) \cdot \mathrm{var}(N))$, a reduced accuracy parameter for sensor b that takes measurement noise into account, and then proceeds as follows:

Algorithm ALG

- **Update opinion:** $x_a(t+1) = x_a(t) + \frac{\tilde{d}_{ab}(t) \cdot \hat{c}_b(t)}{c_a(t) + \hat{c}_b(t)}$.
- **Update accuracy :** $c_a(t+1) = c_a(t) + \hat{c}_b(t)$.

Fix an independent meeting pattern. First, algorithm **ALG** is designed such that at all times, the opinion is preserved as an unbiased estimator of τ^* and the accuracy, $c_a(t)$, remains equal to the reciprocal of the current variance of the opinion $X_a(t, \mathbf{ALG})$. That is, we have:

Lemma 2. *At any round t and for any sensor a: (1) the opinion* $X_a(t, \mathbf{ALG})$ *serves as an unbiased estimator of τ^*, and (2)* $c_a(t) = 1/\mathrm{var}(X_a(t, \mathbf{ALG}))$.

We are now ready to analyze the competitiveness of algorithm **ALG**, by relating the variance of a sensor a at round t to the corresponding *FI*, namely, $J_a(t)$. Recall that Lemma 1 gives a lower bound on the variance of algorithm **Opt** at a sensor a, which depends on the corresponding *FI* at the sensor. Specifically, we have: $\mathrm{var}(X_a(t, \mathbf{Opt})) \geq 1/J_a(t)$. Initially, $J_a(0)$, the *FI* at a sensor a, equals the Fisher information in the parameterized family $\Phi_a(x - \tau)$ with respect to τ, and hence is at most the initial accuracy $c_a(0)$ times Δ_0. We show that the gain in accuracy following an interaction is always at least as large the corresponding upper bound on the gain in Fisher information as given in Theorem 4, divided by the initial Fisher-tightness. That is: $c_a(t+1)-c_a(t) \geq \left(1/(\frac{1}{J_b(t)}+\frac{1}{J_N})\right)/\Delta_0$. Informally, this property of **ALG** can be interpreted as maximizing the Fisher information flow in each interaction up to an approximation factor of Δ_0. By induction, we obtain the following.

Lemma 3. *At every round t, we have $c_a(t) \geq J_a(t)/\Delta_0$.*

Lemmas 1, 2 and 3 can now be combined to yield the following inequality: $\text{var}(X_a(t, \mathbf{ALG})) \leq \Delta_0 \cdot \text{var}(X_a(t, \mathbf{Opt}))$. This establishes Theorem 2. □

Note that if $|F| = O(1)$ (i.e., F contains a constant number of distributions, independent of the number of sensors) then initial Fisher-tightness Δ_0 is a constant, and hence Theorem 2 states that **ALG** is constant-competitive at any sensor and at any time. In some other natural cases the performances of **ALG** are even better. One such case is when the distributions in $\mathcal{F}$ as well as the noise distribution $N(\eta)$ are all Gaussians. In this case $\Delta_0 = 1$ and Theorem 2 therefore states that the variance of **ALG** equals that of **Opt**, for any sensor at at any time. Another case is when $|F|$ is a constant, the noise is Gaussian, and both the population size n and the round t go to infinity. In this case, the performances of **ALG** become arbitrarily close to those of **Opt**.

5 The Fisher Channel Capacity and Convergence Times

For a fixed independent meeting pattern, $J_a(t)$, the FI at a sensor a and round t, was defined in Section 3 with respect to algorithm **Opt**. We note that this definition applies to any algorithm A as long as it is sufficiently smooth so that the corresponding Fisher informations are well-defined. This quantity $J_a(t, A)$ would respect the same recursive inequality as state in Theorem 4, that is, we have: $J_a(t+1, A) \leq J_a(t, A) + \frac{1}{\frac{1}{J_b(t,A)}+\frac{1}{J_N}}$. This directly implies the following:

$$J_a(t+1, A) - J_a(t, A) \leq J_N \ . \tag{2}$$

The inequality above sets a bound of J_N for the increase in *FI* per round. In analogy to Channel Capacity as defined by Shannon [7] we term this upper bound as the *Fisher Channel Capacity.*

The restriction on information flow as given by the Fisher Channel Capacity can be translated into lower bounds for convergence time of algorithm **Opt** (and hence also apply for any algorithm). Recall, ρ is the first round when we have more than half of the population satisfying $\text{var}(X_a(t)) < \epsilon^2$. By Lemma 1, a sensor, a, with variance smaller than ϵ^2 must have a large *FI*, specifically, $J_a(\rho) \geq 1/\epsilon^2$. To get some intuition on the convergence time, assume that the number of sensors is odd, and let J_0 denote the median initial *FI* of sensors (this is the median of the *FI*, J_{Φ_a}, over all sensors a), and assume $J_0 \ll 1/\epsilon^2$. By definition, more than a half of the population have initial Fisher information at most J_0. By the Pigeon-hole principle, at least one sensor has an *FI* of, at most, J_0 at $t = 0$ and, at least, $1/\epsilon^2$ at $t = \rho$. Theorem 1 follows by the fact that, by Equation 2, this sensor could increase its *FI* by, at most, J_N in each observation.

References

1. Angluin, D.: Local and global properties in networks of processors. In: STOC, pp. 82–93 (1980)
2. Attiya, H., Herzberg, A., Rajsbaum, S.: Optimal Clock Synchronization under Different Delay Assumptions. SIAM J. Comput. 25(2), 369–389 (1996)
3. Bar-Yossef, Z., Jayram, T.S., Kumar, R., Sivakumar, D.: Info. Theory Methods in Comm. Complexity. IEEE Conf. on Computational Complexity, 93–102 (2002)
4. Biaz, S., Welch, J.L.: Closed form bounds for clock synchronization under simple uncertainty assumptions. Inf. Process. Lett. 80(3), 151–157 (2001)
5. Blachman, N.M.: The convolution inequality for entropy powers. IEEE Transactions on Information Theory 11(2), 267–271 (1965)
6. Chaudhari, Q., Serpedin, E., Wu, Y.C.: Improved estimation of clock offset in sensor networks. In: ICC (2009)
7. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. John Wiley & Sons (2006)
8. Cristian, F.: Probabilistic Clock Synchronization. Distributed Computing 3(3), 146–158 (1989)
9. Dolev, D., Halpern, J., Simons, B., Strong, R.: Dynamic fault-tolerant clock synchronization. Journal of the ACM 42(1), 143–185 (1995)
10. Elson, J., Girod, L., Estrin, D.: Fine-Grained Network Time Synchronization Using Reference Broadcasts. Operating Systems Review 36, 147–163 (2002)
11. Feinerman, O., Haeupler, B., Korman, A.: Breathe before speaking: efficient information dissemination despite noisy, limited and anonymous communication. In: PODC, pp. 114–123 (2014)
12. Feinerman, O., Korman, A.: Clock Synchronization and Estimation in Highly Dynamic Networks: An Information Theoretic Approach (An Arxiv version). `http://arxiv.org/pdf/1504.08247v1.pdf`
13. El Gamal, A., Kim, Y.: Network Information Theory, 709 p. Cambridge University Press (2012)
14. Gubner, J.: Distributed Estimation and Quantization. IEEE Tran. on Information Theory 39(4) (1993)
15. Jeske, D.: On the maximum likelihood estimation of clock offset. IEEE Trans. Commun. 53(1) (2005)
16. Kar, S., Moura, J.M.F.: Distributed Consensus Algorithms in Sensor Networks With Imperfect Communication: Link Failures and Channel Noise. IEEE Tran. on SIgnal Processing 57(1), 355–369 (2009)
17. Kempe, D., Dobra, A., Gehrke, J.: Gossip-based computation of aggregate information. In: FOCS 2003, pp. 482–449 (2003)
18. Koetter, R., Kschischang, F.R.: Coding for errors and erasures in random network coding. IEEE Transactions on Info. Theory 54(8), 3579–3591 (2008)
19. Korman, A., Greenwald, E., Feinerman, O.: Confidence Sharing: an Economic Strategy for Efficient Information Flows in Animal Groups. PLOS Computational Biology 10(10) (2014)

20. Kuhn, F., Lenzen, C., Locher, T., Oshman, R.: Optimal gradient clock synchronization in dynamic networks. In: PODC, pp. 430–439 (2010)
21. Lamport, L.: Proving the Correctness of Multiprocess Programs. IEEE Transactions on Software Engineering (2), 125–143 (1977)
22. Lenzen, C., Locher, T., Wattenhofer, R.: Tight Bounds for Clock Synchronization. JACM 57(2) (2010)
23. Lenzen, C., Sommer, P., Wattenhofer, R.: PulseSync: An Efficient and Scalable Clock Synchronization Protocol. ACM/IEEE Transactions on Networking (2014)
24. Lenzen, C., Locher, T., Sommer, P., Wattenhofer, R.: Clock synchronization: Open problems in theory and practice. In: van Leeuwen, J., Muscholl, A., Peleg, D., Pokorný, J., Rumpe, B. (eds.) SOFSEM 2010. LNCS, vol. 5901, pp. 61–70. Springer, Heidelberg (2010)
25. Lundelius, J., Lynch, N.: An Upper and Lower Bound for Clock Synchronization. Information and Control 62, 190–204 (1984)
26. McNamara, J.M., Houston, A.I.: Memory and the efficient use of information. Journal of Theoretical Biology 125(4), 385–395 (1987)
27. Mills, D.L.: Internet time synchronization: the network time protocol. IEEE Transactions of Communications 39(10), 1482–1493 (1991)
28. Mills, D.L.: Improved algorithms for synchronizing computer network clocks. Networks 3, 3 (1995)
29. Ostrovsky, R., Patt-Shamir, B.: Optimal and efficient clock synchronization under drifting clocks. In: PODC 1999, pp. 3–12 (1999)
30. Patt-Shamir, B., Rajsbaum, S.: A theory of clock synchronization. In: STOC 1994, pp. 810–819 (1994)
31. Rioul, O.: Information theoretic proofs of entropy power inequalities. IEEE Transactions on Information Theory 57(1), 33–55 (2011)
32. Sivrikaya, F., Yener, B.: Time synchronization in sensor networks: a survey. IEEE Network 18(4) (2004)
33. Shannon, C.: A Mathematical Theory of Communication. Technical Journal 27(3), 379–423 (1948)
34. Solis, R., Borkar, V., Kumar, P.R.: A new distributed time synchronization protocol for multihop wireless networks. In: Proc. 45th IEEE Conference on Decision and Control (CDC) (2006)
35. Stam, A.J.: Some inequalities satisfied by the quantities of information of Fisher and Shannon. Inform. and Control 2, 101–112 (1959)
36. Xiao, L., Boyd, S., Lall, S.: A scheme for robust distributed sensor fusion based on average consensus. In: Proc. of the 4th International Symposium on Information Processing in Sensor Networks (IPSN) (2005)
37. Sundararaman, B., Buy, U., Kshemkalyani, A.D.: Clock synchronization for wireless sensor networks: a survey. Ad Hoc Networks 3, 281–323 (2005)
38. Viswanathan, R., Varshney, P.K.: Distributed detection with multiple sensors I. Fundamentals. Proceedings of the IEEE (1997)
39. Wu, Y.C., Chaudhari, Q.M., Serpedin, E.: Clock Synchronization of Wireless Sensor Networks. IEEE Signal Process. Mag. 28(1), 124–138 (2011)
40. Zamir, R.: A proof of the Fisher Information inequality via a data processing arguement. IEEE Trans. Inf. Theory, 482–491 (2003)

Springer

http://www.springer.com/978-3-319-25257-5

Structural Information and Communication Complexity
22nd International Colloquium, SIROCCO 2015,
Montserrat, Spain, July 14-16, 2015. Post-Proceedings
Scheideler, C. (Ed.)
2015, XI, 476 p. 52 illus., Softcover
ISBN: 978-3-319-25257-5