Error Bounds and Consistency of Maximum Likelihood Time Synchronization

Björn Scheuermann*, Wolfgang Kiess*, Magnus Roos*, Florian Jarre‡, Martin Mauve*

* Heinrich Heine University, Computer Science Department, Düsseldorf, Germany
‡ Heinrich Heine University, Mathematical Department, Düsseldorf, Germany

TECHNICAL REPORT TR-2008-001
HEINRICH HEINE UNIVERSITY, DÜSSELDORF, GERMANY
COMPUTER SCIENCE DEPARTMENT

11 FEBRUARY 2008

Error Bounds and Consistency of Maximum Likelihood Time Synchronization

Björn Scheuermann*, Wolfgang Kiess*, Magnus Roos*, Florian Jarre‡, and Martin Mauve*

*Heinrich Heine University, Computer Science Department, Düsseldorf, Germany
‡Heinrich Heine University, Mathematical Department, Düsseldorf, Germany

Abstract

In this paper we consider a simplified version of a maximum likelihood time synchronization estimator introduced in [2]. This simplified variant assumes approximately accurate clock rates. We give error bounds for the estimator under the assumption of a maximum timestamping delay, and we show that these error bounds are optimal. Furthermore, we prove consistency of the clock offset estimator for increasing numbers of observed events, and we derive a result on the error distribution of the event time estimates.

I. Introduction

In [2] we propose a solution to the problem of generating a time-consistent global log file out of a set of local log files from a number of network nodes. The local clocks of these nodes are inaccurate and thus introduce errors in the local timestamps. In our approach we consider a network with a local broadcast medium, i. e., one where all or parts of the nodes observe certain events at the same time instant. These common events are used as a kind of anchor points for a maximum likelihood estimation of the clock error and of additional delays occuring in each node upon logging the events. The latter are called the timestamping delays. This maximum likelihood estimator leads to a big linear program (LP).

Our proposed approach is able to estimate and compensate linear-affine clock deviations. Here we consider a simplified variant that ignores deviations of the clocks' rates, such that the only deviations of the clocks are constant offsets. We prove two desirable properties for this version of the estimator. First, we show that tight error bounds on the estimation error hold under the assumption of a bounded timestamping delay. In particular this means that the algorithm does not amplify errors. Furthermore, we show that our estimator is consistent. This means that for increasing data set sizes the estimate converges (in probability) to the true values of the estimated features. It thus supports the intuition that the estimate improves for a larger amount of observed and logged events in the nodes.

In this paper we will first quickly review our estimator in Section II, with a special focus on the simplified version without clock rate deviations. We will then focus on the error bounds in Section III. In Section IV we prove the consistency of the estimator. Finally, Section V concludes the paper.

II. The Simplified Maximum Likelihood Time Synchronization Estimator

For the most part we adopt the notation from [2] here. The set of nodes is denoted by J and the set of events by I. R is the relation of node-event pairs for which a timestamp has been recorded, i. e., $(i,j) \in R \subseteq I \times J$ if and only if i has been observed and timestamped by j.

The employed clock model assumes that for each node j there is an offset o_j and a rate r_j such that j's clock maps the real time t to the local time $C_j(t) = r_j t + o_j$. In addition to the clock deviations there are timestamping errors which are assumed to be independent and exponentially distributed with parameter λ. This means that for the full-featured maximum likelihood estimator the time recorded for event i at "real" time T_i by node j is $r_j(T_i + d_{i,j}) + o_j$ where $d_{i,j}$ is a random variable modelling the timestamping delay.

In this paper we consider a simplified version of the above. We assume that the clocks run (approximately) at the correct rate, i. e., we set $\forall j \in J : r_j = 1$. Under this assumption the recorded time for a node-event pair $(i,j) \in R$ becomes $T_i + d_{i,j} + o_j$. Thus the simplified maximum likelihood estimator, in analogy to the full-featured version, is the solution to the following problem:

$$\text{minimize } L = \prod_{(i,j)\in R} \lambda e^{-\lambda\left(t_{i,j} - \widehat{o}_j - \widehat{T}_i\right)}$$

$$\text{subject to } \forall (i,j) \in R : \widehat{d}_{i,j} = t_{i,j} - \widehat{o}_j - \widehat{T}_i \geq 0.$$

Here, $t_{i,j}$ denotes the local time when node j has recoded event i. $\widehat{o}_j$ and $\widehat{T}_i$ are the estimates of o_j and T_i, respectively.

As shown in [2], the optimal solution is independent of λ and equivalent to solving

$$\text{minimize } k(L) = \sum_{(i,j)\in R} \left(t_{i,j} - \widehat{o}_j - \widehat{T}_i\right) = \sum_{(i,j)\in R} \widehat{d}_{i,j}$$

under the same constraints as above.

Let an optimal solution of this LP, consisting of the estimates $\widehat{T}_i$, $\widehat{o}_j$ and $\widehat{d}_{i,j}$, be denoted by $\mathscr{S}$.

To simplify the rest of this paper we introduce notations for the set of all nodes observing a certain event and for all the events observed by a given node:

$$\forall i \in I : R_i := \{j \in J | (i,j) \in R\}$$

$$\forall j \in J : R^j := \{i \in I | (i,j) \in R\}$$

III. Error Bounds

Our intention in this section is to establish an upper bound on the error of the maximum likelihood estimator, i. e., the maximum difference between estimated and real event times. In order to do so we make two additional assumptions. The first one guarantees network connectivity, the second one establishes an upper bound on the timestamping delay.

Before we introduce the assumptions it is necessary to recall one important fact from [2], which we call the *offset ambiguity*: for any offset $\tau \in \mathbb{R}$, the estimates $\widehat{T}_i$, $\widehat{o}_j$ and the estimates $\widehat{T}_i'$, $\widehat{o}_j'$ with

$$\forall i \in I : \widehat{T}_i' = \widehat{T}_i + \tau \qquad\qquad \forall j \in J : \widehat{o}_j' = \widehat{o}_j - \tau$$

fit the same set of measurements equally well.

From the offset ambiguity it is easy to see that there is also no way to estimate all the relative times within an experiment if the network is partitioned. If there are no anchor points between two sets of nodes, there will be an ambiguity of the offset between these partitions within the experiment. Thus, in order to get a bounded maximum estimation error we need to assume network connectivity. The *connectivity assumption* says that the network nodes do not fall into disjoint partitions, between which no anchor points at all exist. Formally this is

$$\forall P_1, P_2 \subseteq J,\ P_1, P_2 \neq \emptyset : (P_1 \cup P_2 = J \ \wedge\ P_1 \cap P_2 = \emptyset \quad \Longrightarrow \quad \exists j_1 \in P_1, j_2 \in P_2 : R^{j_1} \cap R^{j_2} \neq \emptyset).$$

Under this assumption we will prove that

$$\forall j_1, j_2 \in J : \left|(o_{j_1} - o_{j_2}) - (\widehat{o}_{j_1} - \widehat{o}_{j_2})\right| \leq (|J| - 1) \cdot D$$

and

$$\forall i_1, i_2 \in I : \left|(T_{i_1} - T_{i_2}) - \left(\widehat{T}_{i_1} - \widehat{T}_{i_2}\right)\right| \leq |J| \cdot D$$

if $D \in \mathbb{R}^+$ is an upper bound for the delays, i. e.,

$$\forall (i, j) \in R : d_{i,j} \leq D.$$

Note that the bounds are on the difference between two estimation errors because of the offset ambiguity.

The following proof does not exploit the exponential distribution of the delays. Thus, independent of the derivation of the estimator, the proof shows that if there is an upper bound for the timestamping delays the estimates are close to the real values, regardless of the real distribution of the delays within $[0, D]$. Note that assuming the existence of such an upper bound does not limit the practical applicability of the results given here: for any given experiment, the set of observations R is finite, and thus there will always be a maximum delay.

For the proof we introduce some additional terminology.

Definition Let $j, j' \in J, i \in I$. We call i a *common event* of j and j' iff $\{(i, j_1), (i, j_2)\} \subseteq R$, and we call i a *connecting event* from j to j' iff i is a common event of j and j' and the timestamping delay of i in j is estimated as zero, i. e., $\widehat{d_{i,j}} = 0$.

As will soon become clear, a significant part of the timestamping delay estimates are zero. We now construct a directed graph $G := (J, E)$ with

$$E := \left\{(j, j') \in J^2 \mid \exists i \in I : i \text{ is a connecting event from } j \text{ to } j'\right\}.$$

The graph G plays a central role in our proof. In the following two lemmas we point out some properties of G.

Lemma III.1 *Let $j_A, j_B \in J$. Then there exists a directed path $(j_A, \dots, j_B)$ in G.*

Proof The set of nodes J can be divided into two disjoint subsets:

$$\begin{aligned} J_1 &:= \{j \in J \mid \text{there exists no directed path } (j, \dots, j_B) \text{ in } G\} \\ J_2 &:= J \setminus J_1. \end{aligned}$$

In the following, we will show that J_1 is empty. Let $\bar{I}$ be the events occurring in both J_1 and J_2:

$$\bar{I} := \{i \in I \mid \exists j_1 \in J_1, j_2 \in J_2 : \{(i, j_1), (i, j_2)\} \subseteq R\}.$$

Let $j_1 \in J_1, j_2 \in J_2$. Then there is no connecting event from j_1 to j_2. Otherwise a path from j_1 to j_B could be constructed by concatenation of (j_1, j_2) and $(j_2, \dots, j_B)$, which is a contradiction to $j_1 \in J_1$. Thus, we have

$$\exists \varepsilon > 0 : \forall (i, j) \in R \cap (\bar{I} \times J_1) : \widehat{d}_{i,j} \geq \varepsilon.$$

Let now

$$\begin{aligned} I_1 &:= \{i \in I \mid \nexists j \in J_2 : (i, j) \in R\} \\ I_2 &:= \{i \in I \mid \nexists j \in J_1 : (i, j) \in R\}. \end{aligned}$$

With these definitions $I = I_1 \cup I_2 \cup \bar{I}$ holds, and $I_1, I_2, \bar{I}$ are pairwise disjoint. Now a new solution $\mathscr{S}'$ of the LP can be constructed:

$$\begin{aligned} \widehat{T}_i' &:= \widehat{T}_i - \begin{cases} \varepsilon & \text{if } i \in I_1 \\ 0 & \text{otherwise} \end{cases} \\ \widehat{o}_j' &:= \widehat{o}_j + \begin{cases} \varepsilon & \text{if } j \in J_1 \\ 0 & \text{otherwise} \end{cases} \\ \widehat{d}_{i,j}' &:= \widehat{d}_{i,j} - \begin{cases} \varepsilon & \text{if } i \in \bar{I} \ \wedge \ j \in J_1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

It can easily be verified that all the LP constraints hold for $\mathscr{S}'$ since they hold for $\mathscr{S}$. For $\mathscr{S}'$ we have

$$\sum_{(i,j) \in R} \widehat{d}_{i,j}' = \sum_{(i,j) \in R} \widehat{d}_{i,j} - |R \cap (\bar{I} \times J_1)| \cdot \varepsilon.$$

Since $\mathscr{S}$ is optimal and $\varepsilon > 0$ we have $|R \cap (\bar{I} \times J_1)| = 0$. Therefore, $\bar{I} = \emptyset$ or $J_1 = \emptyset$, due to the definitions of J_1 and $\bar{I}$ above.

$\bar{I}$ is not empty if J_1 is not empty due to the connectivity assumption. Therefore, $J_1 = \emptyset$ and $j_A \in J_2$. Thus, there exists a directed path $(j_A, \ldots, j_B)$ in G. ∎

Lemma III.2 *Let $(j_0, \ldots, j_n) \in J^{n+1}$ be a directed path in G. Then*

$$(o_{j_0} - o_{j_n}) - (\widehat{o}_{j_0} - \widehat{o}_{j_n}) \leq nD.$$

Proof We prove this by induction.

Let $n = 1$. Let i be a connecting event from j_0 to j_1. Such an event exists due to the construction of G. From the LP we have

$$\begin{aligned} T_i + o_{j_0} + d_{i,j_0} &= \widehat{T}_i + \widehat{o}_{j_0} + \widehat{d}_{i,j_0} && (1) \\ T_i + o_{j_1} + d_{i,j_1} &= \widehat{T}_i + \widehat{o}_{j_1} + \widehat{d}_{i,j_1}. && (2) \end{aligned}$$

The difference between (1) and (2) yields

$$\begin{aligned} & (o_{j_0} - o_{j_1}) + (d_{i,j_0} - d_{i,j_1}) &=& \ (\widehat{o}_{j_0} - \widehat{o}_{j_1}) + (\widehat{d}_{i,j_0} - \widehat{d}_{i,j_1}) \\ \iff \quad & (o_{j_0} - o_{j_1}) - (\widehat{o}_{j_0} - \widehat{o}_{j_1}) &=& \ \underbrace{(\underbrace{\widehat{d}_{i,j_0}}_{=0} - \underbrace{\widehat{d}_{i,j_1}}_{\geq 0})}_{\leq 0} - \underbrace{(\underbrace{d_{i,j_0}}_{\in[0,D]} - \underbrace{d_{i,j_1}}_{\in[0,D]})}_{\in[-D,D]} \\ \iff \quad & (o_{j_0} - o_{j_1}) - (\widehat{o}_{j_0} - \widehat{o}_{j_1}) &\leq& \ D. \end{aligned}$$

For the induction step we have:

$$\begin{aligned} (o_{j_0} - o_{j_{n-1}}) - (\widehat{o}_{j_0} - \widehat{o}_{j_{n-1}}) &\leq (n-1)D && (3) \\ (o_{j_{n-1}} - o_{j_n}) - (\widehat{o}_{j_{n-1}} - \widehat{o}_{j_n}) &\leq D. && (4) \end{aligned}$$

(4) can be constructed like above. Addition of (3) and (4) gives us

$$\begin{aligned} (o_{j_0} - o_{j_{n-1}}) - (\widehat{o}_{j_0} - \widehat{o}_{j_{n-1}}) + (o_{j_{n-1}} - o_{j_n}) - (\widehat{o}_{j_{n-1}} - \widehat{o}_{j_n}) &\leq (n-1)D + D \\ \iff \quad (o_{j_0} - o_{j_n}) - (\widehat{o}_{j_0} - \widehat{o}_{j_n}) &\leq nD. \end{aligned}$$

∎

Now that we know that G is connected and that an upper error bound holds for any path in G, we only need to put together these pieces in order to get a bounded error of the offsets for any pair of nodes. While Lemma III.2 establishes only an upper bound for the differences of two offset estimation errors, the fact that G is connected and thus paths in both directions exist can be exploited to bound this difference from below, too.

Theorem III.3 *Let $j_1, j_2 \in J$. Then the following bound holds*

$$\left|(o_{j_1} - o_{j_2}) - (\widehat{o}_{j_1} - \widehat{o}_{j_2})\right| \leq (|J| - 1)D.$$

Proof According to Lemma III.1 there exists a path from j_1 to j_2 in G. Likewise, there exists a path from j_2 to j_1 in G. Since the number of nodes in G is $|J|$, the maximum length of each of these paths is $|J|-1$. Thus, we have from Lemma III.2:

$$(o_{j_1}-o_{j_2})-(\widehat{o}_{j_1}-\widehat{o}_{j_2}) \leq (|J|-1)\,D$$
$$(o_{j_2}-o_{j_1})-(\widehat{o}_{j_2}-\widehat{o}_{j_1}) = -((o_{j_1}-o_{j_2})-(\widehat{o}_{j_1}-\widehat{o}_{j_2})) \leq (|J|-1)\,D.$$

This immediately gives us the desired result. ∎

We now have an error bound for the offset estimates. From here, it is only a small step to a similar bound for the error in the estimation of the event times. However, in order to give a good bound we need an additional property of the estimator that will be established in the following lemma.

Lemma III.4 *Let $\bar{i} \in I$. Then there exists $\bar{j} \in J : \widehat{d}_{\bar{i},\bar{j}} = 0$.*

Proof We prove this by contradiction. Assume that there is no such $\bar{j}$. Then

$$\exists \varepsilon > 0 : \forall (i,j) \in R : i = \bar{i} \Rightarrow \widehat{d}_{i,j} \geq \varepsilon.$$

Now we construct an new solution $\mathscr{S}'$ to the LP:

$$\begin{aligned}
\widehat{T}_i' &:= \widehat{T}_i + \begin{cases} \varepsilon & \text{if } i = \bar{i} \\ 0 & \text{otherwise} \end{cases} \\
\widehat{o}_j' &:= \widehat{o}_j \\
\widehat{d}_{i,j}' &:= \widehat{d}_{i,j} - \begin{cases} \varepsilon & \text{if } i = \bar{i} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

This solution is valid and better than $\mathscr{S}$. Since $\mathscr{S}$ is optimal this is a contradiction. ∎

We can exploit the fact that for any event i a node j exists where $d_{i,j}$ is estimated as zero in the following theorem. The result is the desired error bound for the event time estimates.

Theorem III.5 *Let $i_1, i_2 \in I$. Then the following bound holds*

$$\left|(T_{i_1}-T_{i_2})-(\widehat{T}_{i_1}-\widehat{T}_{i_2})\right| \leq |J|\cdot D.$$

Proof From Lemma III.4 we have

$$\forall k \in \{1,2\} : \exists j_k \in J : (i_k, j_k) \in R \;\wedge\; \widehat{d}_{i_k,j_k} = 0.$$

With these j_1, j_2 we have

$$T_{i_1} + o_{j_1} + d_{i_1,j_1} = \widehat{T}_{i_1} + \widehat{o}_{j_1} \tag{5}$$
$$T_{i_2} + o_{j_2} + d_{i_2,j_2} = \widehat{T}_{i_2} + \widehat{o}_{j_2}. \tag{6}$$

Calculating the difference between (5) and (6) and some reordering yields

$$(T_{i_1} - T_{i_2}) - (\widehat{T}_{i_1} - \widehat{T}_{i_2}) = \underbrace{(\widehat{o}_{j_1} - \widehat{o}_{j_2}) - (o_{j_1} - o_{j_2})}_{\in[-(|J|-1)D,\ (|J|-1)D]} - \underbrace{(d_{i_1,j_1} - d_{i_2,j_2})}_{\in[-D,D]}.$$

The bounds for $(\widehat{o}_{j_1} - \widehat{o}_{j_2}) - (o_{j_1} - o_{j_2})$ come from theorem III.3.

This result gives us

$$\left|(T_{i_1} - T_{i_2}) - (\widehat{T}_{i_1} - \widehat{T}_{i_2})\right| \leq (|J|-1)D + D = |J| \cdot D,$$

which is the desired bound. ∎

Now we have bounds on the relative offsets and event time estimates. It remains open whether these bounds are good. To address this question we introduce some additional terms. We then use these to point out an important property of any local broadcast network where distributed log files are recorded. This property in turn will then be used to prove that not only the error bounds are tight, but also that no estimator with better error bounds is possible.

Definition A *receive trace set* $Q = (I, R, (t_{i,j})_{(i,j)\in R})$ for a set of nodes J consists of a set of events I, a relation $R \subseteq I \times J$ and the local timestamps $t_{i,j}$ recorded by the nodes in J for the events they have observed.

For a given set of nodes J, a *scenario* $S = (I, R, (o_j)_{j\in J}, (T_i)_{i\in I}, (d_{i,j})_{(i,j)\in R})$ consists of a set of events I, a relation $R \subseteq I \times J$, and corresponding offsets o_j for all nodes, event times T_i for all events and delays $d_{i,j}$ for each event reception.

From the above definition it is clear that there is exactly one receive trace set for any given scenario. There can, however, well be many possible scenarios for a given receive trace set. The offset ambiguity mentioned at the beginning of this section is a special case of this situation. Note that this is an inherent property of the time synchronization problem as it is considered here, and not specific to our approach.

Definition Two scenarios $S = (I, R, (o_j)_{j\in J}, (T_i)_{i\in I}, (d_{i,j})_{(i,j)\in R})$ and $S' = (I, R, (o'_j)_{j\in J}, (T'_i)_{i\in I}, (d'_{i,j})_{(i,j)\in R})$ are called *indistinguishable* if they share a common set of events I and the same relation R and they result in the same receive trace set, i. e.,

$$\forall (i,j) \in R : o_j + T_i + d_{i,j} = t_{i,j} = o'_j + T'_i + d'_{i,j}.$$

Theorem III.6 *For any set of nodes J, $|J| \geq 1$ there exist two indistinguishable scenarios $S = (I, R, (o_j)_{j\in J}, (T_i)_{i\in I}, (d_{i,j})_{(i,j)\in R})$ and $S' = (I, R, (o'_j)_{j\in J}, (T'_i)_{i\in I}, (d'_{i,j})_{(i,j)\in R})$ and $j_2, j_2 \in J$ such that R fulfills the connectivity assumption and*

$$(o'_{j_1} - o'_{j_2}) - (o_{j_1} - o_{j_2}) = 2(|J|-1)D.$$

Proof Our proof for this theorem is constructive. Let, for the sake of simplicity and without loss of generality, $J = \{1, \ldots, n\}$. Assume these nodes form a chain-like topology. A total of $n-1$ events are recorded by the nodes, $I = \{1, \ldots, n-1\}$. Now we define R as follows

$$(i, j) \in R \quad \Longleftrightarrow \quad i \in \{j-1, j\}.$$

It is easy to see that the connectivity assumption holds for R.

Now let S and S' be defined by

$$\begin{aligned}
\forall i \in I: \qquad & T_i := (n-i) \cdot D & T_i' := i \cdot D \\
\forall j \in J: \qquad & o_j := j \cdot D & o_j' := (n-j+1) \cdot D \\
\forall (i,j) \in R: \qquad & d_{i,j} := \begin{cases} 0 & \text{if } i = j-1 \\ D & \text{if } i = j \end{cases} & d_{i,j}' := \begin{cases} D & \text{if } i = j-1 \\ 0 & \text{if } i = j. \end{cases}
\end{aligned}$$

With these definitions it is easy to verify that S and S' are indistinguishable:

$$\forall (i,j) \in R: T_i + o_j + d_{i,j} = T_i' + o_j' + d_{i,j}'.$$

And it also holds that

$$(o_1' - o_n') - (o_1 - o_n) = 2(n-1)D.$$

$\blacksquare$

Let us now assume that we have two scenarios S and S' like in the above theorem. Assume now we have estimates $\widehat{o}_1$ and $\widehat{o}_n$ of o_1 and o_n that are better than the worst-case result of our maximum likelihood estimator for scenario S. This means that the relative offset estimation error is less than $(|J|-1)D$, which is the bound of our approach according to Theorem III.3. Then in particular

$$(o_1 - o_n) - (\widehat{o}_1 - \widehat{o}_n) > -(|J|-1)D.$$

But we know that

$$(o_1' - o_n') - (o_1 - o_n) = 2(n-1)D.$$

By simple addition we get

$$(o_1' - o_n') - (\widehat{o}_1 - \widehat{o}_n) > (|J|-1)D.$$

This is worse than the maximum error of our MLE estimator. Since S and S' are indistinguishable, this proves that there cannot be an estimator with lower maximum offset estimation errors: if the estimate is better in case of a receive trace set that resulted from S, it is necessarily worse in case the same receive trace set came from S' and vice versa. Similar results hold for the event time error bounds.

IV. Consistency

In the last section error bounds for our time synchronization approach have been established. They do not exploit the exponential distribution of the timestamping delay, but rely on an upper bound for the timestamping delay. In this section we will not assume such an upper bound, but we will exploit the exponential distribution of the delays. Under these premises consistency of the clock offset estimator will be established, which means convergence in probability to the correct offset values for an increasing number of observed events:

$$\forall j \in J : \operatorname*{plim}_{|I|\to\infty} \widehat{o}_j = o_j + x,$$

where $x \in \mathbb{R}$ again comes from the offset ambiguity discussed in the previous section.

We will show the consistency of the simplified MLE under an additional regularity condition, defined as follows.

Definition We say that the *regularity condition* is fulfilled if there exists an undirected, connected graph $G = (J,V)$ and some positive constant β such that

$$\forall \{j_1, j_2\} \in V : E\left[\left|\{i \in I | \{j_1, j_2\} \subseteq R_i\}\right|\right] \geq \beta \cdot |I|.$$

It is valid to assume that G is a tree.

This condition can be seen as a somewhat stronger variant of the connectivity assumption used in the previous section. It is stronger in the sense that it requires an ever-growing number of independent connections between all parts of the network with an increasing total number of observed events, although only probabilistically with respect to the expectancy of their number.

A. Stochastic preliminaries

Before we can tackle the main proof some preliminary results from elementary probability theory are necessary. They will be established in the following lemmas.

Lemma IV.1 *Let X_1, X_2 be independent, exponentially distributed random variables with parameters λ_1, λ_2. Then $\min\{X_1, X_2\}$ is exponentially distributed with parameter $\lambda_1 + \lambda_2$.*

Proof

$$\begin{aligned}
P(\min\{X_1,X_2\} \leq x) &= P(X_1 \leq x \text{ or } X_2 \leq x) \\
&= 1 - P(X_1 > x \text{ and } X_2 > x) \\
&= 1 - P(X_1 > x) \cdot P(X_2 > x) \\
&= 1 - \left(\int_x^\infty \lambda_1 e^{-\lambda_1 t} dt\right) \left(\int_x^\infty \lambda_2 e^{-\lambda_2 t} dt\right) \\
&= 1 - \left(e^{-\lambda_1 x}\right) \left(e^{-\lambda_2 x}\right) \\
&= 1 - e^{-(\lambda_1+\lambda_2)x}.
\end{aligned}$$

Thus, the minimum of X_1 and X_2 is exponentially distributed with parameter $\lambda_1 + \lambda_2$. ∎

Lemma IV.2 *Let X be a random variable with real values and expected value $E[X] \in \mathbb{R}$. Let $t \in \mathbb{R}$ such that $E[X|X < t]$ and $E[X|X \geq t]$ exist. Then*

$$E[X] = P(X < t) \cdot E[X|X < t] + P(X \geq t) \cdot E[X|X \geq t].$$

Proof Assume X is continuous with probability density f.

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f(x)\ dx \\ &= \int_{-\infty}^{t} x \cdot f(x)\ dx + \int_{t}^{\infty} x \cdot f(x)\ dx \\ &= \int_{-\infty}^{t} x \cdot f(x|X < t) \cdot P(X < t)\ dx \\ &+ \int_{t}^{\infty} x \cdot f(x|X \geq t) \cdot P(X \geq t)\ dx \\ &= P(X < t) \cdot E[X|X < t] + P(X \geq t) \cdot E[X|X \geq t]. \end{aligned}$$

For more general X the proof is analogous. ∎

Lemma IV.3 *Let d be an exponentially distributed random variable with parameter λ. Let $t \in \mathbb{R}^+$. Then*

$$E[d|d < t] = \frac{1}{\lambda} - \frac{te^{-\lambda t}}{1 - e^{-\lambda t}}.$$

Proof From Lemma IV.2 and with the memorylessness of the exponential distribution it follows that

$$\begin{aligned} E[d|d < t] &= \frac{E[d] - P(d \geq t) \cdot E[d|d \geq t]}{P(d < t)} \\ &= \frac{\frac{1}{\lambda} - e^{-\lambda t}\left(\frac{1}{\lambda} + t\right)}{1 - e^{-\lambda t}} \\ &= \frac{(1 - e^{-\lambda t})\frac{1}{\lambda} - te^{-\lambda t}}{1 - e^{-\lambda t}} \\ &= \frac{1}{\lambda} - \frac{te^{-\lambda t}}{1 - e^{-\lambda t}}. \end{aligned}$$

∎

Lemma IV.4 *Let $d_1, \ldots, d_n$ be independent, exponentially distributed random variables with parameters $\lambda_1, \ldots, \lambda_n$. Let $\Delta_1, \ldots, \Delta_n \in \mathbb{R}$ be given such that $\forall i, 1 \leq i < n : \Delta_{i+1} \geq \Delta_i$.*

With

$$\forall i, 1 \leq i \leq n : L_i := \sum_{j=1}^{i} \lambda_j$$

the following equality holds:

$$E\left[\min_{1\leq i\leq n}(d_i + \Delta_i)\right] = \sum_{j=1}^{n-1}\left(\prod_{k=1}^{j-1} e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\left(1 - e^{-L_j(\Delta_{j+1}-\Delta_j)}\right)\frac{1}{L_j}$$
$$+\left(\prod_{k=1}^{n-1} e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\frac{1}{L_n} + \Delta_1.$$

Proof First observe that if there exists $i, 1 \leq i < n$ such that $\Delta_{i+1} = \Delta_i$, then, with Lemma IV.1, this is equivalent to the case where d_i and d_{i+1} are replaced by a single exponentially distributed random variable with parameter $\lambda_i + \lambda_{i+1}$, and the same Δ_i appearing only once. We may thus without loss of generality assume that $\forall i, 1 \leq i < n : \Delta_{i+1} > \Delta_i$.

To simplify the notation of the following, we define

$$\chi\left[(\Delta_1, \lambda_1), \ldots, (\Delta_n, \lambda_n)\right] := E\left[\min_{1\leq j\leq n}(d_j + \Delta_j)\right].$$

This can be generalized for the conditional expected value, i. e., for a condition A, let

$$\chi[(\Delta_1, \lambda_1), \ldots, (\Delta_n, \lambda_n)|A] := E\left[\min_{1\leq j\leq n}(d_j + \Delta_j)\middle|A\right].$$

We will now show the assertion by induction over n. The base case $n = 1$ holds:

$$\chi[(\Delta_1, \lambda_1)] = \Delta_1 + \frac{1}{\lambda_1}.$$

For the induction step, we will use the following implication of Lemma IV.1:

$$\begin{aligned}
&\chi[(\Delta_2, \lambda_1), (\Delta_2, \lambda_2), (\Delta_3, \lambda_3), \ldots, (\Delta_n, \lambda_n)] \\
&\qquad = E\left[\min\{d_1 + \Delta_2, d_2 + \Delta_2, d_3 + \Delta_3, \ldots, d_n + \Delta_n\}\right] \\
&\qquad = E\left[\min\{\min\{d_1 + \Delta_2, d_2 + \Delta_2\}, d_3 + \Delta_3, \ldots, d_n + \Delta_n\}\right] \\
&\qquad = \chi[(\Delta_2, \lambda_1 + \lambda_2), (\Delta_3, \lambda_3), \ldots, (\Delta_n, \lambda_n)].
\end{aligned}$$

With $\Delta_1 \leq \Delta_2$, Lemma IV.2 at (a), the memorylessness of the exponential distribution at (b), Lemma IV.3 at (c) and the induction hypothesis (with an index shift) at (d), the following holds:

$$
\begin{aligned}
&\chi[(\Delta_1,\lambda_1),\ldots,(\Delta_{n+1},\lambda_{n+1})]\\
&= E\left[\min\{d_1+\Delta_1,\ldots,d_{n+1}+\Delta_{n+1}\}\right]\\
&\overset{(a)}{=} P(d_1+\Delta_1<\Delta_2)\cdot E\left[\min\{d_1+\Delta_1,\ldots,d_{n+1}+\Delta_{n+1}\} \mid d_1+\Delta_1<\Delta_2\right]\\
&\quad + P(d_1+\Delta_1\geq\Delta_2)\cdot E\left[\min\{d_1+\Delta_1,\ldots,d_{n+1}+\Delta_{n+1}\} \mid d_1+\Delta_1\geq\Delta_2\right]\\
&\overset{(b)}{=} P(d_1+\Delta_1<\Delta_2)\cdot E\left[d_1+\Delta_1 \mid d_1+\Delta_1<\Delta_2\right]\\
&\quad + P(d_1+\Delta_1\geq\Delta_2)\cdot E\left[\min\{d_1+\Delta_2,d_2+\Delta_2,\ldots,d_{n+1}+\Delta_{n+1}\}\right]\\
&= \left(1-e^{-\lambda_1(\Delta_2-\Delta_1)}\right)\left(E[d_1 \mid d_1<\Delta_2-\Delta_1]+\Delta_1\right)\\
&\quad + e^{-\lambda_1(\Delta_2-\Delta_1)}\cdot\chi[(\Delta_2,\lambda_1),(\Delta_2,\lambda_2),\ldots,(\Delta_{n+1},\lambda_{n+1})]\\
&\overset{(c)}{=} \left(1-e^{-\lambda_1(\Delta_2-\Delta_1)}\right)\left(\frac{1}{\lambda_1}-\frac{e^{-\lambda_1(\Delta_2-\Delta_1)}(\Delta_2-\Delta_1)}{1-e^{-\lambda_1(\Delta_2-\Delta_1)}}+\Delta_1\right)\\
&\quad + e^{-\lambda_1(\Delta_2-\Delta_1)}\cdot\chi[(\Delta_2,\lambda_1+\lambda_2),(\Delta_3,\lambda_3),\ldots,(\Delta_{n+1},\lambda_{n+1})]\\
&\overset{(d)}{=} \left(1-e^{-\lambda_1(\Delta_2-\Delta_1)}\right)\left(\frac{1}{\lambda_1}+\Delta_1\right)-e^{-\lambda_1(\Delta_2-\Delta_1)}(\Delta_2-\Delta_1)\\
&\quad + e^{-\lambda_1(\Delta_2-\Delta_1)}\left(\sum_{j=2}^{n}\left(\prod_{k=2}^{j-1}e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\left(1-e^{-L_j(\Delta_{j+1}-\Delta_j)}\right)\frac{1}{L_j}\right.\\
&\quad \left. + \left(\prod_{k=2}^{n}e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\frac{1}{L_{n+1}}+\Delta_2\right)\\
&= \left(1-e^{-L_1(\Delta_2-\Delta_1)}\right)\left(\frac{1}{L_1}+\Delta_1\right)-e^{-L_1(\Delta_2-\Delta_1)}(\Delta_2-\Delta_1)\\
&\quad + e^{-L_1(\Delta_2-\Delta_1)}\left(\sum_{j=2}^{n}\left(\prod_{k=2}^{j-1}e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\left(1-e^{-L_j(\Delta_{j+1}-\Delta_j)}\right)\frac{1}{L_j}\right.\\
&\quad \left. + \left(\prod_{k=2}^{n}e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\frac{1}{L_{n+1}}+\Delta_2\right)\\
&= \Delta_1+\left(1-e^{-L_1(\Delta_2-\Delta_1)}\right)\frac{1}{L_1}\\
&\quad + \sum_{j=2}^{n}\left(\prod_{k=1}^{j-1}e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\left(1-e^{-L_j(\Delta_{j+1}-\Delta_j)}\right)\frac{1}{L_j}\\
&\quad + \left(\prod_{k=1}^{n}e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\frac{1}{L_{n+1}}\\
&= \Delta_1+\sum_{j=1}^{n}\left(\prod_{k=1}^{j-1}e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\left(1-e^{-L_j(\Delta_{j+1}-\Delta_j)}\right)\frac{1}{L_j}\\
&\quad + \left(\prod_{k=1}^{n}e^{-L_k(\Delta_{k+1}-\Delta_k)}\right)\frac{1}{L_{n+1}}.
\end{aligned}
$$

∎

B. *Consistency proof*

We will now give the main consistency proof of the simplified maximum likelihood time synchronization estimator. In order to do so, we first formalize the notion of the clock offset estimation error. This definition is actually very trivial; we denote the estimation error by a vector $\Delta = (\Delta_1, \ldots, \Delta_{|J|})^T \in \mathbb{R}^{|J|}$ in the following way.

Definition Let $\forall j \in J : \Delta_j := o_j - \widehat{o}_j$.

In the following, we regard a certain scenario (according to the definition in Section III) as given, thus the T_i, o_j and $d_{i,j}$ are fixed. We then consider the estimation error vector Δ as variable and have a look at the properties of the likelihood function upon a varying estimation error. Our goal is to show that the probability that the likelihood function (regarded as a function of Δ) has its optimum in an arbitrarily small environment around the correct clock offset estimates is arbitrarily high for a sufficing number of observed events.

On our way towards this goal we now introduce a per-event decomposition of the objective function $k(L)$. Certain properties of these event-wise objective function terms form the basis of our proof.

Definition For each event $i \in I$ let f_i be the term added to the objective function $k(L)$ by i for some estimation error vector Δ:

$$f_i(\Delta) := \sum_{j \in R_i} \widehat{d}_{i,j}.$$

The constraints of the LP in our approach are of the form

$$\forall (i,j) \in R : \widehat{d}_{i,j} + \widehat{o}_j + \widehat{T}_i = t_{i,j} = d_{i,j} + o_j + T_i.$$

Thus, the estimated timestamping delays can be expressed as

$$\begin{aligned} \widehat{d}_{i,j} &= d_{i,j} + o_j - \widehat{o}_j + T_i - \widehat{T}_i \\ &= d_{i,j} + \Delta_j + (T_i - \widehat{T}_i). \end{aligned}$$

Considering that the $d_{i,j}$ and the T_i are given by the scenario it is easy to see that the Δ_j determine the value of the event time estimates $\widehat{T}_i$ chosen by the estimator: all the estimated delays $\widehat{d}_{i,j}$ need to be non-negative and at the same time the sum of all these delay estimates is minimized. Thus, the optimal choice is

$$\widehat{T}_i = T_i + \min_{k \in R_i} (d_{i,k} + \Delta_k).$$

This also follows from the nonnegativity constraints for the $\widehat{d}_{i,j}$ and Lemma III.4.

We can now express the objective function terms f_i in the following way:

$$f_i(\Delta) = \sum_{j \in R_i} \left(d_{i,j} + \Delta_j - \min_{k \in R_i} \left(d_{i,k} + \Delta_k \right) \right).$$

This is the formulation that we are going to use throughout the proof. Now we will point out some properties of the objective function terms.

Lemma IV.5 *The objective function terms $f_i(\Delta)$ are convex.*

Proof It is well-known that the minimum of concave functions is concave (see, e. g., [1]). Hence, $\min_{k\in R_i}(d_{i,k}+\Delta_k)$ is concave and $-\min_{k\in R_i}(d_{i,k}+\Delta_k)$ is convex. As the sum of convex functions is convex the assertion follows. ∎

The next property that we are going to prove is a little more tricky. For simplicity, we will assume that $J=\{1,\dots,|J|\}$ from now on.

Lemma IV.6 *There exists a strictly monotonically increasing function $\alpha:\mathbb{R}_0^+\to\mathbb{R}_0^+$ such that for each event $i\in I$ and each $\Delta\in\mathbb{R}^{|J|}$ the following holds:*

$$E[f_i(\Delta)] \quad \geq \quad E[f_i(0)] \quad + \quad \alpha(\max_{\substack{(j_1,j_2)\in R_i\times R_i \\ \nexists k\in R_i:\Delta_{j_1}<\Delta_k<\Delta_{j_2}}}(\Delta_{j_2}-\Delta_{j_1})).$$

Proof Let $n:=|R_i|$. By the linearity of the expected value, $E[f_i(\Delta)]$ can be rewritten in the following way:

$$\begin{aligned}
E[f_i(\Delta)] =& E\left[\sum_{j\in R_i}\left(d_{i,j}+\Delta_j-\min_{k\in R_i}(d_{i,k}+\Delta_k)\right)\right] \\
=& E\left[\sum_{j\in R_i}d_{i,j}\right]+E\left[\sum_{j\in R_i}\Delta_j\right]-n\cdot E\left[\min_{j\in R_i}(d_{i,j}+\Delta_j)\right] \\
=& \frac{n}{\lambda}+\sum_{j\in R_i}\Delta_j-n\cdot E\left[\min_{j\in R_i}(d_{i,j}+\Delta_j)\right].
\end{aligned}$$

Assume without loss of generality that $R_i=\{1,\dots,n\}$ and that the nodes are ordered such that $\forall j, 1\leq j<n:\Delta_{j+1}\geq\Delta_j$. The exponential distributions if the $d_{i,j}$ all share the same parameter λ, thus with Lemma IV.4 we have

$$\begin{aligned}
E[f_i(\Delta)] =& \frac{n}{\lambda}+\sum_{j=1}^{n}\Delta_j-n\Delta_1 \\
& -n\cdot\sum_{j=1}^{n-1}\left(\prod_{k=1}^{j-1}e^{-k\lambda(\Delta_{k+1}-\Delta_k)}\right)\left(1-e^{-j\lambda(\Delta_{j+1}-\Delta_j)}\right)\frac{1}{j\lambda} \\
& -n\cdot\left(\prod_{k=1}^{n-1}e^{-k\lambda(\Delta_{k+1}-\Delta_k)}\right)\frac{1}{n\lambda}.
\end{aligned}$$

We now define $\forall j, 1\leq j<n:\delta_j:=\Delta_{j+1}-\Delta_j$ and $\delta:=(\delta_1,\dots,\delta_{n-1})^T$. Then $\forall j, 1\leq j<n:$ $\delta_j\geq 0$. (Note that δ has dimension zero for $n=1$.) With the given definition of δ the following reformulation of $E[f_i(\Delta)]$ is possible:

$$\begin{aligned}
E[f_i(\Delta)] =& \frac{n}{\lambda}+\sum_{j=1}^{n-1}(n-j)\delta_j \\
& -\sum_{j=1}^{n-1}\left(\prod_{k=1}^{j-1}e^{-k\lambda\delta_k}\right)\left(1-e^{-j\lambda\delta_j}\right)\frac{n}{j\lambda}-\left(\prod_{k=1}^{n-1}e^{-k\lambda\delta_k}\right)\frac{1}{\lambda}.
\end{aligned}$$

We generalize this to a function K of arbitrary vectors θ with non-negative elements and dimension $m := \dim\theta$, i. e.

$$K(m,\theta) := \frac{m+1}{\lambda} + \sum_{j=1}^{m}(m-j+1)\theta_j - \sum_{j=1}^{m}\left(\prod_{k=1}^{j-1} e^{-k\lambda\theta_k}\right)\left(1-e^{-j\lambda\theta_j}\right)\frac{m+1}{j\lambda} - \left(\prod_{k=1}^{m} e^{-k\lambda\theta_k}\right)\frac{1}{\lambda}.$$

Note that

$$E[f_i(\Delta)] = K(n-1,\delta).$$

Now we construct α. Let e_k^d be the k-th unit vector of dimension d and set

$$\alpha(t) := \min_{\substack{1\le d<|J|\\ 1\le k\le d}} \left(K(d,t\cdot e_k^d) - K(d,0)\right).$$

For $n=1$ the assertion is true since for $n=1$

$$\max_{\substack{(j_1,j_2)\in R_i\times R_i\\ \nexists k\in R_i:\Delta_{j_1}<\Delta_k<\Delta_{j_2}}} (\Delta_{j_2}-\Delta_{j_1}) = 0$$

and $\alpha(0)=0$.

We thus focus on the case $n>1$ now. Note that in this case

$$\max_{\substack{(j_1,j_2)\in R_i\times R_i\\ \nexists k\in R_i:\Delta_{j_1}<\Delta_k<\Delta_{j_2}}} (\Delta_{j_2}-\Delta_{j_1}) = ||\delta||_\infty.$$

For each m, K is differentiable in δ. We will now show that for all $m\ge 1$ and all p with $1\le p\le m$

$$\frac{\partial K(m,\theta)}{\partial\theta_p} > 0$$

for $\theta_p>0$. Calculating the partial derivative explicitly yields

$$\begin{aligned}\frac{\partial K(\theta)}{\partial\theta_p} &= (m-p+1) + p\sum_{j=p+1}^{m}\left(\prod_{k=1}^{j-1} e^{-k\lambda\theta_k}\right)\left(1-e^{-j\lambda\theta_j}\right)\frac{m+1}{j} \\ &\quad - \left(\prod_{k=1}^{p} e^{-k\lambda\theta_k}\right)(m+1) + p\left(\prod_{k=1}^{m} e^{-k\lambda\theta_k}\right) \\ &= m-p+1 \\ &\quad + p\left(\prod_{k=1}^{p} e^{-k\lambda\theta_k}\right)\left(\sum_{j=p+1}^{m}\left(\prod_{k=p+1}^{j-1} e^{-k\lambda\theta_k}\right)\left(1-e^{-j\lambda\theta_j}\right)\frac{m+1}{j}\right. \\ &\quad \left. - \frac{m+1}{p} + \left(\prod_{k=p+1}^{m} e^{-k\lambda\theta_k}\right)\right).\end{aligned}$$

Since $\frac{m+1}{j} > 1$ for $1 \leq j \leq m$ and

$$\sum_{j=p+1}^{m}\left(\prod_{k=p+1}^{j-1} e^{-k\lambda\theta_k}\right)\left(1-e^{-j\lambda\theta_j}\right)+\left(\prod_{k=p+1}^{m} e^{-k\lambda\theta_k}\right)=1$$

the following holds:

$$\begin{aligned}\frac{\partial K(\theta)}{\partial\theta_p} &\geq m-p+1+p\underbrace{\left(\prod_{k=1}^{p} e^{-k\lambda\theta_k}\right)}_{\leq 1}\underbrace{\left(1-\frac{m+1}{p}\right)}_{<0}\\ &\geq m-p+1+p\left(1-\frac{m+1}{p}\right)\\ &=0.\end{aligned}$$

Here, the first inequality is strict if there is a $j > p$ such that $\theta_j > 0$, and the second inequality is strict if there is a $j \leq p$ such that $\theta_j > 0$.

Now the inequality from the assertion is easily verified. Let r ($1 \leq r < n$) be an index for which $\delta_r = ||\delta||_\infty$. Such an r exists by the definition of $||\cdot||_\infty$. Then

$$\begin{aligned} & E[f_i(\Delta)]-E[f_i(0)]\\ = \;& K(|R_i|,\delta)-K(|R_i|,0)\\ \geq \;& K(|R_i|,||\delta||_\infty\cdot e_r^{|R_i|})-K(|R_i|,0)\\ \geq \;& \min_{1\leq k\leq|R_i|}\left(K(|R_i|,||\delta||_\infty\cdot e_k^{|R_i|})-K(|R_i|,0)\right)\\ \geq \;& \alpha(||\delta||_\infty).\end{aligned}$$

The first inequality stems from the fact that all entries of the Jacobian of K are non-negative and δ is component-wise greater than $||\delta||_\infty\cdot e_r^{|R_i|}$.

It remains to show that $\alpha(||\delta||_\infty) > 0$ is strictly monotonically increasing. This, however, is easy to see. From the above calculations we know that the entries of the Jacobian of $K(m,\theta)$ are *strictly* positive if there is some k such that $\theta_k > 0$. $\alpha(t)$ is the minimum of

$$K(d,t\cdot e_k^d)-K(d,0)$$

for a finite number of combinations of d and k. For each of these combinations and for any $t_1 > t_2 \geq 0$, though, $(K(d,t_1\cdot e_k^d)-K(d,0)) > (K(d,t_2\cdot e_k^d)-K(d,0))$, since the k-th component of $t_1\cdot e_k^d$ is strictly greater than the respective component of $t_2\cdot e_k^d$, whereas all other components are equal. Thus, $\alpha(t_1) > \alpha(t_2)$ and the assertion holds. ∎

Lemma IV.7 *There is some* $\mathscr{L}\in\mathbb{R}$ *such that*

$$\frac{1}{|I|}\cdot k(L)=\frac{1}{|I|}\sum_{(i,j)\in R}\widehat{d}_{i,j}$$

is Lipschitz continuous in Δ *with Lipschitz constant* $\mathscr{L}$. $\mathscr{L}$ *does not depend on* $|I|$.

Proof From the closed form of f_i given above it can be seen that f_i is continuous in Δ for each event $i \in I$. It is also easy to see that the partial derivatives of $f_i(\Delta)$ exist almost everywhere and are, where they exist, bounded above by 1 and below by $-|R_i|$, and thus also by $-|J|$. Thus, all f_i are Lipschitz continuous with a common Lipschitz constant $\mathscr{L}$.

Since

$$k(L) = \sum_{i \in I} f_i(\Delta)$$

we can conclude that $k(L)$ is Lipschitz continuous with Lipschitz constant $|I| \cdot \mathscr{L}$. Therefore, $\mathscr{L}$ is also a Lipschitz constant for $\frac{1}{|I|} \cdot k(L)$. ∎

From the definition of f_i above it can be seen that $f_i(\Delta) = f_i(\Delta + (t, \ldots, t)^T)$ for any $t \in \mathbb{R}$. This is again the offset ambiguity. Thus, from now on we can assume without loss of generality that the estimation error for the $|J|$-th node is zero. Therefore, we will ignore this node in the following and reduce Δ to a vector of dimension $|J| - 1$. Consistency of the MLE is then equivalent to

$$\operatorname*{plim}_{|I| \to \infty} ||\Delta|| = 0.$$

In the following theorem we will use infinity norm spheres. Our notation for them is as follows.

Definition

$$\forall m \in \mathbb{R}^n, r \in \mathbb{R} : S_\infty(m, r) := \left\{ x \mid ||x - m||_\infty < r \right\}$$
$$\forall m \in \mathbb{R}^n, r \in \mathbb{R} : \bar{S}_\infty(m, r) := \left\{ x \mid ||x - m||_\infty = r \right\}$$

Theorem IV.8 *If the regularity condition is fulfilled, then for all $\varepsilon, \delta > 0$ there exists an $N \in \mathbb{N}$ such that from $|I| \geq N$ it follows that for*

$$\tilde{\Delta} := \operatorname*{argmin}_{\Delta \in \mathbb{R}^j} \sum_{i \in I} f_i(\Delta)$$

the following holds

$$P(||\tilde{\Delta}||_\infty > \delta) < \varepsilon$$

i. e.,

$$\operatorname*{plim}_{|I| \to \infty} ||\Delta|| = 0.$$

Thus, the simplified MLE is consistent.

Proof Let

$$r := \frac{\beta \cdot \alpha \left(\frac{\delta}{(|J|-1)^2} \right)}{3 \cdot \mathscr{L}}$$

with β from the regularity condition, α from Lemma IV.6 and $\mathscr{L}$ from Lemma IV.7.

Let $M \subset \bar{S}_\infty(0, \delta) \subset \mathbb{R}^{|J|-1}$ be a finite set of points such that

$$\bar{S}_\infty(0, \delta) \subset \bigcup_{m \in M} S_\infty(m, r).$$

Such a set of points exists since $\bar{S}_\infty(0,\delta)$ is compact and $r > 0$. Let m be one of the points in M. Then there exists $p \in J$ such that $m_p = \delta$. Let $G = (J,V)$ be the graph from the regularity condition. Then there is a path from p to the node with ID $|J|$ with a length of at most $|J|-1$. The estimation error of the $|J|$-th node is 0 by assumption. Thus, there exists $\{j_1, j_2\} \in V$ such that $|m_{j_1} - m_{j_2}| \geq \frac{\delta}{|J|-1}$.

Let i be an event for which $\{j_1, j_2\} \subseteq R_i$. Since the total number of nodes is $|J|$, there are at most $|J|$ nodes in R_i, and for two of them, j_1 and j_2, it holds that $|m_{j_1} - m_{j_2}| \geq \frac{\delta}{|J|-1}$. Thus,

$$\max_{\substack{(j_1',j_2') \in R_i \times R_i \\ \nexists k \in R_i : m_{j_1'} < m_k < m_{j_2'}}} (m_{j_2'} - m_{j_1'}) \quad \geq \quad \frac{\delta}{(|J|-1)^2}.$$

We define

$$I^+ := \{i \in I | \{j_1, j_2\} \subseteq R_i\}.$$

From the regularity condition we know that

$$E[|I^+|] \geq \beta \cdot |I|.$$

This yields

$$\begin{aligned}
& E\left[\frac{1}{|I|}\sum_{i\in I} f_i(m)\right] \\
= \ & \frac{1}{|I|}\left(E\left[\sum_{i\in I\setminus I^+} f_i(m)\right] + E\left[\sum_{i\in I^+} f_i(m)\right]\right) \\
\geq \ & \frac{1}{|I|}\left(E\left[\sum_{i\in I\setminus I^+} f_i(0)\right] + E\left[\sum_{i\in I^+}\left(f_i(0) + \alpha\left(\frac{\delta}{(|J|-1)^2}\right)\right)\right]\right) \\
= \ & \frac{1}{|I|}\left(E\left[\sum_{i\in I\setminus I^+} f_i(0)\right] + E\left[\sum_{i\in I^+} f_i(0) + |I^+| \cdot \alpha\left(\frac{\delta}{(|J|-1)^2}\right)\right]\right) \\
= \ & \frac{1}{|I|}\left(E\left[\sum_{i\in I^+} f_i(0)\right] + E\left[\sum_{i\in I\setminus I^+} f_i(0)\right] + E[|I^+|] \cdot \alpha\left(\frac{\delta}{(|J|-1)^2}\right)\right) \\
\geq \ & \frac{1}{|I|}\left(E\left[\sum_{i\in I} f_i(0)\right] + \beta \cdot |I| \cdot \alpha\left(\frac{\delta}{(|J|-1)^2}\right)\right) \\
= \ & E\left[\frac{1}{|I|}\sum_{i\in I} f_i(0)\right] + \beta \cdot \alpha\left(\frac{\delta}{(|J|-1)^2}\right).
\end{aligned}$$

From the Lipschitz condition established in Lemma IV.7 and the definition of r above we get that for all $m' \in \mathbb{R}^{|J|-1}$ with $||m'-m|| < r$ the following holds:

$$\left| \frac{1}{|I|} \sum_{i \in I} f_i(m) - \frac{1}{|I|} \sum_{i \in I} f_i(m') \right| < r \cdot \mathscr{L} = \frac{1}{3} \cdot \beta \cdot \alpha \left(\frac{\delta}{(|J|-1)^2} \right).$$

Thus, we can conclude that if

$$\frac{1}{|I|} \sum_{i \in I} f_i(m) > E\left[\frac{1}{|I|} \sum_{i \in I} f_i(0) \right] + \frac{2}{3} \cdot \beta \cdot \alpha \left(\frac{\delta}{(|J|-1)^2} \right) \tag{7}$$

holds for m, then for all $m' \in S_\infty(m, r)$

$$\frac{1}{|I|} \sum_{i \in I} f_i(m') > E\left[\frac{1}{|I|} \sum_{i \in I} f_i(0) \right] + \frac{1}{3} \cdot \beta \cdot \alpha \left(\frac{\delta}{(|J|-1)^2} \right). \tag{8}$$

By the law of large numbers there is some $N_m \in \mathbb{N}$ for m where for $|I| \geq N_m$ it follows that

$$P\left(\frac{1}{|I|} \sum_{i \in I} f_i(m) > E\left[\frac{1}{|I|} \sum_{i \in I} f_i(0) \right] + \frac{2}{3} \cdot \beta \cdot \alpha \left(\frac{\delta}{(|J|-1)^2} \right) \right) \geq 1 - \frac{\varepsilon}{|M|+1}.$$

We call the condition in the probability above the superiority condition for m. Similar to the superiority conditions, there is an inferiority condition: there exists some $N_0 \in \mathbb{N}$ such that for $|I| \geq N_0$

$$P\left(\frac{1}{|I|} \sum_{i \in I} f_i(0) < E\left[\frac{1}{|I|} \sum_{i \in I} f_i(0) \right] + \frac{1}{3} \cdot \beta \cdot \alpha \left(\frac{\delta}{(|J|-1)^2} \right) \right) \geq 1 - \frac{\varepsilon}{|M|+1}.$$

Because $|M|$ is finite, there is some $N^* = \max(\{N_0\} \cup \{N_m | m \in M\})$ fulfilling the superiority conditions for all $m \in M$ as well as the inferiority condition, each with a probability of at least $1 - \frac{\varepsilon}{|M|+1}$. Thus, for $|I| \geq N^*$ the probability that all $|M|+1$ conditions are fulfilled is at least

$$1 - (|M|+1) \cdot \frac{\varepsilon}{|M|+1} = 1 - \varepsilon.$$

Since the spheres $S_\infty(m, r)$ around all $m \in M$ cover $\bar{S}_\infty(0, \delta)$, (8) holds for all $m' \in \bar{S}_\infty(0, \delta)$ if (7) holds for all $m \in M$. Hence, with probability of at least $1 - \varepsilon$, it holds that

$$\forall m' \in \bar{S}_\infty(0, \delta) : \frac{1}{|I|} \sum_{i \in I} f_i(m') > \frac{1}{|I|} \sum_{i \in I} f_i(0).$$

Therefore, we know that $\frac{1}{|I|} \sum_{i \in I} f_i$ has a local optimum in $S_\infty(0, \delta)$. Since f_i is convex for each i by Lemma IV.5, $\frac{1}{|I|} \sum_{i \in I} f_i$ is also convex. Thus, the local optimum is also a global optimum. A global optimum of $\frac{1}{|I|} \sum_{i \in I} f_i$ is also a global optimum of $\sum_{i \in I} f_i$. Therefore, for the vector $\tilde{\Delta}$ from the optimal LP solution we have

$$P(\tilde{\Delta} \in S_\infty(0, \delta)) = P(||\tilde{\Delta}|| \leq \delta) \geq 1 - \varepsilon.$$

This is the assertion. ∎

From the consistency result regarding the clock offsets it is easy to obtain a result on the quality of the event time estimates in the same asymptotic scenario. It has been stated before that the estimates of the event times are given by $\widehat{T_i} = T_i + \min_{k \in R_i}(d_{i,k} + \Delta_k)$. Thus, if Δ is close to zero (neglecting the offset ambiguity), the estimate for event i will be wrong by $\min_{k \in R_i} d_{i,k}$. From Lemma IV.1 it then follows that the error is exponentially distributed with parameter $|R_i| \cdot \lambda$. In particular this means that—as could be expected—the expected estimation error decreases with the number of nodes observing the same event.

V. Conclusion

In this paper we have considered a simplified variant of the maximum likelihood time synchronization estimator presented in [2]. The results presented here underline the performance and the practical applicability of our time synchronization approach.

We have first proven tight error bounds for the estimates under the assumption of bounded maximum timestamping delays in the network nodes. It has furthermore been shown that the estimator is optimal in the sense that no estimator with a smaller maximum clock offset estimation error is possible. Although exponentially distributed timestamping delays are a major building block in the derivation of the estimator, the validity of the bounds does not depend on this assumption.

Moreover, by exploiting the exponential distribution assumption of the timestamping delays we have given a proof for the consistency of the clock offset estimates. Therefrom we have also derived a result on the error distribution of the event time estimates.

References

[1] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Springer, Berlin–Heidelberg–New York, 1991.

[2] B. Scheuermann, W. Kiess, M. Roos, F. Jarre, and M. Mauve. On the Time Synchronization of Distributed Log Files in Networks with Local Broadcast Media. *IEEE/ACM Transactions on Networking*, 2008. Accepted for publication.